

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2003, 09:32:34 ; Search time 2444 Seconds
(without alignments)
4366.950 Million cell updates/sec

Title: US-09-802-285A-2
Perfect score: 3494
Sequence: 1 MTTFKRIIVFAVIALSSG.....KGLNLTITNGKQQLVLP 659

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US09802285/runat_28072003_190511_9730/app_query.fasta_1.839
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09802285@cgn_1_1906 @runat_28072003_190511_9730 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	109.5	3.1	2547	11	AK010256	AK010256 Mus muscu
2	107.5	3.1	2637	11	AK011146	AK011146 Mus muscu
3	99.5	2.8	957	12	BG566122	BG566122 602582746
C 4	98.5	2.8	580	10	AW942324	LD14783.3
5	97.5	2.8	641	13	BJ420555	BJ420555 BJ420555
6	97.5	2.8	655	13	BJ424970	BJ424970 BJ424970
7	97.5	2.8	659	13	BJ421576	BJ421576 BJ421576
8	97.5	2.8	663	13	BJ411853	BJ411853 BJ411853
9	97.5	2.8	815	14	BQ752776	WHE4119.A
10	95.5	2.7	1050	11	AK006271	AK006271 Mus muscu
11	95	2.7	671	14	BQ744898	946112B02
12	94	2.7	1001	17	CNS06JUZ	AL401685 T7 end of
13	93	2.7	756	13	BM413117	BM413117 EST587444
14	93	2.7	762	13	BM165623	BM165623 EST568146
15	93	2.7	780	13	BI647560	BI647560 603278139
16	93	2.7	953	12	BF686272	BF686272 602142733
C 17	92.5	2.6	625	9	AL649582	AL649582
18	92.5	2.6	716	13	BM161568	BM161568 EST564091
19	92.5	2.6	734	13	BI889777	BI889777 ZF637-2-0
C 20	92.5	2.6	966	12	BG120749	BG120749 602348507
21	92.5	2.6	2587	11	AK010578	AK010578 Mus muscu
22	92	2.6	809	9	AL669439	AL669439
23	92	2.6	2588	11	AY108732	AY108732 Zea mays
24	91.5	2.6	953	14	BQ891785	BQ891785 AGENCOURT
C 25	91.5	2.6	966	12	BG114631	BG114631 602315418
26	91.5	2.6	1072	10	BE378651	BE378651 601237043
C 27	91	2.6	329	17	AZ769390	IM0570H03
28	91	2.6	611	17	BH670936	BH670936 BOHTE55PR
29	91	2.6	668	12	BF294532	BF294532 006PB03
30	91	2.6	883	14	BQ687587	BQ687587 AGENCOURT
31	91	2.6	947	17	CNS015IM	AL105496 Drosophil
C 32	91	2.6	1028	17	CNS048YA	AL279883 Tetraodon
33	90.5	2.6	568	17	BH371376	BH371376 AG-ND-125
34	90.5	2.6	584	14	BQ450998	BQ450998 PFEST0a0
35	90.5	2.6	743	14	BQ917010	BQ917010 QHB20A01
C 36	90.5	2.6	1200	14	BQ890483	BQ890483 AGENCOURT
37	90	2.6	621	13	BJ338430	BJ338430
38	90	2.6	713	9	AU101720	AU101720 AU101720
39	90	2.6	875	17	BH155081	BH155081 ENT0223TF
40	90	2.6	929	13	BI147055	BI147055 602913334
C 41	90	2.6	1036	17	CNS07B0G	AL437270 T3 end of
42	90	2.6	1896	11	AK003169	AK003169 Mus muscu
43	89.5	2.6	586	13	BM317522	BM317522 LG1_332.F
44	89.5	2.6	625	17	AQ034246	Cpg0543A
45	89.5	2.6	658	14	BQ519807	rd01g12.y

ALIGNMENTS

RESULT 1
AK010256
LOCUS
DEFINITION
Mus musculus ES cells cDNA, RIKEN full-length enriched library,
clone:2400002G22:DNA segment, Chr 15, Wayne State University 59,
expressed, full insert sequence.
ACCESSION
AK010256
VERSION
AK010256.1 GI:12845561
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (strain:C57BL/6J) ES cells cDNA to mRNA,
clone_lib:RIKEN full-length enriched mouse cDNA library
clone:2400002G22.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

[illegible][illegible]

Query Match:	3.08%	Indels:	220
DB:	11	Gaps:	40
US-09-802-285A-2 (1-659) x AK011146 (1-2637)			
QY	18	SerSerGlyAsnIleLeuAla-GlnSerSerSerIleThrArgLysAspPheAspHisI1	37
DB	561	AGCCAGCGCGTATACCTGACCTCCATGATGAGATGAAATCAAAGGTTGTTGGATTT	1562
QY	37	eAsnLeuGluTyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAs	57
DB	621	C-----TATGGCGAGTCTATGTTAGCGGCCACCAATAGAAAACGGCTTCTATTA	671
QY	57	pAspAlaAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAs	77
DB	672	TGAT-----ATGTACTTTGAAGAAGGGGGCGTGTCCAGCAATGA	710
QY	77	pPheSerAsnAlaGlu-----LysProAlaAspIleArgGlnProIleAspLys	93
DB	711	CTTCTCTCTCCCTGGAACCTTTGTGCAAGAAATCATTTAAAGAGAAACAACTTTGCAAG	770
QY	93	sValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTy	113
DB	771	ATTG-----GAAGTTAAAGGAACATTTACTAGAAATGTTTAAG-----TA	812
QY	113	rGlyTyrPheAsp-----TyrGlyLysAspIleAsn-----Tr	124
DB	813	CAACAAGTTCAGTCCGGATATTGAATGAGAAAGTGAATACTCCGACCAGACTGTCTA	872
QY	124	pGlnMetTrpProValLysAsp-----AsnGluValArg-----	135
DB	873	CAGATGTGGCCCTGTAGATCTCTCCGGGGTCTCATGTCGACACACCGGCAAGAT	932
QY	136	-----TrpGlnLeuHisArg-----ValLysTrpTrpGlnAlaMetAla	148
DB	933	TAAACATTGAAGATACACAGAATTTCTTCACGTACTGGGAAGCGAAGCAGACATGGA	992
QY	149	-----LeuValTyrHisAlaThr-----GlyAspGluLysTyrAlaArgGluTrpVa	164
DB	993	GACGCTCCAGCGGATCTATGCAATTTCAATTCCTGACCCCAAGCTACTGAAAGAGTGGA	1052
QY	164	lTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLy	184
DB	1053	GAAGTTCCAGAGGAAGCCAAAACCGAGACACACAGGAAAATCGGAGCGGACCAAGAACT	1112
QY	184	sPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProThrPheSe	204
DB	1113	ATAATTTCTC-----CATGAACATTAGCCCTGGAAGTTG	1145
QY	204	rLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTy	224
DB	1146	CTTTTCTCTGCCAAAGAGGCTACATTTATAATACACTAATGGAATTTATCAGAGT--	1203
QY	224	rHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPh	244
DB	1204	-----GAATATAGAAAAGAGGGTTCCAGGAGGTCGTCTCACTCCCAACATCTT	1250
QY	244	eGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAspSerProAr	264
DB	1251	CAATAGCGCGTCTGGATGACCTCTGGC-----CA	1280
QY	264	gTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleLysLysGlnValTyrAlaAs	284
DB	1281	CTGGCAGCACTACACGAGAACATGTTCTCTCTTTCAGGTGGAGAAGGAG-----	1329
QY	284	pGlyMetGlnPheGluLeuSerProIleTyr-----HisValAlaAlaIleAsp--	300
DB	1330	-----CAGTTTGACATTAAAGCCCATGAACCTGCCCGGACACTGCCTGTATTCGATCA	1382
QY	301	-----IlePheLeuLysAlaTyrGlySerAlaLysAr	311
DB	1383	CGGGCCACGGTCTCTGGCGAGAGCTGCCCTCGGGCTAGCTACTGTTGTGCTGCATAG	1442
QY	311	gValAsnLeuGlu-----LysGluPheProGlnSer--	321
DB	1443	GAATGAGCTCTCGGGGGCTCTCACCGGGCTCACACGGGTCCGAAGATTCACAGAGATGA	1501
QY	322	-----TyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerI1	336
DB	1503	CGCACACATCTCTGTGCCATGGAGCAGATGAAGATGAAATCAAAGGTTGTTGGATTT	1562
QY	336	eSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPh	356
DB	1563	TCCTCCACAGTATATAGT-----GTCITTTGGATTTTCATTTAAATTCGAATCTTCTAC	1616
QY	356	eArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTy	376
DB	1617	TCGCCAGAAAATCTCTGGAGATATTGAAATATGG-----AACCAAGCT-----	1662
QY	376	rPheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAs	396
DB	1663	-----GAGAAACAACCTTGAAAA	1679
QY	396	nAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLy	416
DB	1680	CAGT-----TTCATGAGTTCGGTGAGAAGTGGGAACCTTAAT-----	1716
QY	416	sAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheI1	436
DB	1717	-----CCTGGAGATGGAGCCTTC-----	1734
QY	436	eLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaI1	456
DB	1735	-----TATGGGCCAAAGATTGCATACAGATA-----AAGATGCCAT	1772
QY	456	eMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLysAspAs	476
DB	1773	CGTCCG-----TACCACCACTGT-----GCACCACTCCAGCTGCACIT	1811
QY	476	nGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAs	496
DB	1812	TCAGTTTGCCTATC-----AGGTTTAAC-----	1833
QY	496	pValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPh	516
DB	1834	-----CTTACTTACGTC-----AGCCACCATGCGCATGATAAGAAAGGCCAGTGAT--	1881
QY	516	eIleAsnLysLysTyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLe	536
DB	1882	GTTCACAGAGCCATCTCGGGTCGGTGGNAAGATGATGCCATCTCACAGAAAACCTA	1940
QY	536	uGlyValHisTrpGln-----LeuLysGluAspSerAsnPr	548
DB	1941	TGGGGGCAATGGCTTTCTGGCTCTCTCTGCCAGGTGATGGTGTCCAGTGGGACC	2000
QY	548	oValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGlyAsnAsnLeuMe	568
DB	2001	GACATGTGATGAATATGCCCAAAGGTACGGCAACAATTCATGATGCTAAATTCATGCC	2060
QY	568	tIleGlnSerLeuAsnAlaAspArgThrSerLeuAsnGluGluGluLysValSerTy	588
DB	2061	GGACACTGACCTGGATCCAGGCTGTACC-----TTGAATAGAAG-----	2100
QY	588	rValTyrAsnL	

```

Db      2220 GGAGGAGCGGTGGCGCGCTGCAGCAGCTC 2250
RESULT 3
BG566122
LOCUS   957 bp mRNA linear EST 10-APR-2001
DEFINITION 602582746F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710659 5',
mRNA sequence.
ACCESSION BG566122
VERSION   BG566122.1 GI:13573775
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 957)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI549 row: k column: 12
High quality sequence stop: 739.
Location/Qualifiers
1. 957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4710659"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatgcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCATATATGCC-3', and 3' adaptor sequence:
5'-ATTCTAGGCGGAGCGCGGACATG-3' (30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 269 a 256 c 254 g 178 t
ORIGIN

Alignment Scores:
Pred. No.: 0.294 Length: 957
Score: 99.50 Matches: 64
Percent Similarity: 33.56% Conservative: 33
Best Local Similarity: 22.15% Mismatches: 121
Query Match: 2.85% Indels: 71
DB: 12 Gaps: 13

US-09-802-285A-2 (1-659) x BG566122 (1-957)

QY 187 TrpArgProLeuGluValSerAspArgValGlnSerLeuProThrPheSerLeuPhe 206
Db 185 TTTTGAACAGGAGTAGTAAATAGTCGTTTCAGGAGATCCAAAGCACTCTTTTACTTC 244
QY 207 ValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGln 226
Db 245 ACCCCCAACAGACAGGACACACATTTCTTCAGAGATAC----CAGACCCGACAGAC 301
QY 227 GlnAlaAspTyrLeuSerThrHis-----TyrAlaGluGlnGlyAsnHisArgLeu 243
Db 302 CAGTGCATCTATAACACACCATCTACCTGAATGTCAGCGGGAATGGACCATCTCCAGA 361
QY 244 PheGluAlaGlnArgAsnLeuPheAlaGly----- 253

Db      362 TACGTGGGAGCCCAAGAGCAATTCGCTCACTTGTGATCCTCAGGAGCACCAAGACCTAC 421
QY 254 ---ValSerPheProGluPheLysAspSerProArgTTrpArgGlnThrGlyLeuSerVal 272
Db 422 ATGCTTGTCTTTT---GAGCTGAACGATGAGAAGAACTGG-----GGGCTGTCT--- 466
QY 273 LeuAsnThrGluIleLeuLysGlnValTyrAlaAsp----- 284
Db 467 -----GTCTATCTGACAAAGCCAGACGACGACCAAGAGCAAA 502
QY 285 ---GlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeu 303
Db 503 CTGGGAGAGTTCTACGAGCTCTCGACTGTGTCGTCATTCCTCCCAAGTCAGATGCTGTGT 562
QY 304 LysAlaTyrGlySerAlaLysArgValAsn---LeuGluLysGluPheProGlnSerTyr 322
Db 563 ACACCGATTGGAAAAACGGAATAGTGTGAGCCACTGGAGAAAGCAGACGACGAGGAGAG 622
QY 323 ValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsn 342
Db 623 GAAACCAAGAGGAGGAGGAGGAGTCTAGCCAGGACA----- 658
QY 343 ThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPheAla 362
Db 659 -----CAGCCTTGATCCAGACAGACGACTTGGGGGCCATCTGACCCCTC 703
QY 363 SerTrpAlaArgVal-----PheProAlaAsn-----GlnAlaIle 374
Db 704 CAAACCCGACAAATGTTTACCTCAGCTTTTCCCTTCAGCTTCCAGTAGTAGGCTATC 763
QY 375 LysTyrPheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeu 394
Db 764 CTGTGTTTGGACCAAGAAACAGGCTTCAACACACAAAGAAATGTTTA 823
QY 395 SerAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetVal 414
Db 824 AGCGGCGCT-----TGGGACGCGAGAAAGAGGCTCTTAGTA 859
QY 415 Leu-----LysAlaSerProGlyGluPheHisAlaGlnProAspAsnGly 430
Db 860 CCACACAGGTGGGACCAACGCGCGCCCAAAACATGACAGCAGCAGGAGGAGGAGGCGC 919
QY 431 ThrPheGluLeuPheIleLysGlyArg 439
Db 920 ACCAAGAGAGACCGGTGAAGGAACGG 946

RESULT 4
AW942324/c
LOCUS   580 bp mRNA linear EST 23-APR-2001
DEFINITION LD14783.3prime LD Drosophila melanogaster embryo Bluescript
Drosophila melanogaster cDNA clone LD14783 3, mRNA sequence.
ACCESSION AW942324
VERSION   AW942324.1 GI:8120024
KEYWORDS  EST.
SOURCE    fruit fly.
ORGANISM  Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 580)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
Other ESTs: LD14783.5prime
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Based upon the presence of a XhoI site followed by a run of 14 or

```



```

Db 420 TTAGATCAAGCAATAGTGGTGTATATCTTTTCAATCAAACCTTTTCCCAATTGAT 479
Qy 403 ArgSerGlyTyrAspLysAsnAlaThrValMetValLeuLysAlaSerProProGlyGlu 422
Db 480 AACATGGGGTACGATGTAGATCCATCAATTAGATTATATAACAAAGT-----GGTAAC 533
Qy 423 PheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLys--GlyArgAsnPhe 441
Db 534 TATCAT-----AATTTCCATTTTGTGTAAAGATGAATAGTCGTTTT 575

Qy 442 ThrProAspAlaGly---ValPheValTyrSerGlyAspGlu 454
Db 576 ACTTATAATGGAGGTGAAGTTTATAATTAAAGGTGATGAT 617

RESULT 6
BJ424970 655 bp mRNA linear EST 11-MAR-2002
LOCUS BJ424970 Dictyostelium discoideum cDNA library, VF Dictyostelium
DEFINITION BJ424970 Dictyostelium discoideum cDNA clone ddv54c02 5', mRNA sequence.
ACCESSION BJ424970
VERSION BJ424970.1 GI:19341677
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
AUTHORS 1 (bases 1 to 655)
TITLES Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1..655
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv54c02"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
/dev_stage="Growth phase"
BASE COUNT 250 a 94 c 88 g 223 t
ORIGIN

Alignment Scores:
Pred. No.: 0.289 Length: 655
Score: 97.50 Matches: 49
Percent Similarity: 36.92% Conservative: 30
Best Local Similarity: 22.90% Mismatches: 70
Query Match: 2.79% Indels: 65
DB: 13 Gaps: 12

US-09-802-285a-2 (1-659) x BJ424970 (1-655)

Qy 269 GlyIleSerValLeuAsnThrGluIleLysLys-----GlnValTyrAla 283
Db 99 GGTGTGAGTTTAATGATCACTAATCAATTAATCAAGTTCCAAATGGCCCAACAATTATATT 158
Qy 284 AspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeu 303
Db 159 GAAGGTACAAATTTTGATCAATGGCCAGAATATAAT----- 194
Qy 304 LysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrVal 323
Db 195 -----TATAATTTTGAAAAACCAACCCAGGAGTAAATGTT 230
Qy 324 GlnThrValGluAsnMetIleMetAlaLeuIle-----SerIleSerLeuProAspTyr 341

```

```

Db 231 AAAACCCCTT---GGAAATGTTAAAGCAAAATTTGAATCAACAAACAAAGTTCCAGAACTT 287
Qy 342 AsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPhe 361
Db 288 TCTTCA-----GATGATACTAAAACTTCCACCAATAACATTAGGTACTATGCAATAT 338
Qy 362 AlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAspGly 381
Db 339 CCA-----GAGTTATTATTAATATCACTTTTACATCAAT 371
Qy 382 LysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeu----- 394
Db 372 GTTGATGCACCCAGCATCCAAAACTCTCGTAAAAAATTTACCAATGAAAAATACAATTCAA 431
Qy 395 -----SerAsnAlaGlyPheTyrThrPhe----- 402
Db 432 TTAGATCAAGCAAAATAGTGGTGTATATCTTTTCAATCAAACCTTTTCCCAATTGAT 491
Qy 403 ArgSerGlyTyrAspLysAsnAlaThrValMetValLeuLysAlaSerProProGlyGlu 422
Db 492 AACATGGGCTAGATGTAGATCCATCACTTTAGAAATTTATAACAAAGT-----GGTAAC 545
Qy 423 PheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLys---GlyArgAsnPhe 441
Db 546 TATCAT-----AATTTCCATTTTGTGTAAAGATGAATAGTCGTTTT 587
Qy 442 ThrProAspAlaGly---ValPheValTyrSerGlyAspGlu 454
Db 588 ACTTATAATGGGGTGAAGTTTATAATTAAAGGTGATGAT 629

RESULT 7
BJ421576 659 bp mRNA linear EST 10-MAR-2002
LOCUS BJ421576 Dictyostelium discoideum cDNA library, VF Dictyostelium
DEFINITION BJ421576 Dictyostelium discoideum cDNA clone ddv43k03 5', mRNA sequence.
ACCESSION BJ421576
VERSION BJ421576.1 GI:19333816
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
AUTHORS 1 (bases 1 to 659)
TITLES Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1..659
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv43k03"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
/dev_stage="Growth phase"
BASE COUNT 254 a 94 c 88 g 223 t
ORIGIN

Alignment Scores:
Pred. No.: 0.292 Length: 659
Score: 97.50 Matches: 49
Percent Similarity: 36.92% Conservative: 30
Best Local Similarity: 22.90% Mismatches: 70
Query Match: 2.79% Indels: 65
DB: 13 Gaps: 12

```

```

US-09-802-285A-2 (1-659) x BJ421576 (1-659)
/strain="AX4"
/db_xref="taxon:44689"
/clone="adv6g05"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
/sex="mat A"
/dev_stage="Growth phase"
BASE COUNT 253 a -92 c 90 g 228 t
ORIGIN
Alignment Scores:
Pred. No.: 0.295 Length: 663
Score: 97.50 Matches: 49
Percent Similarity: 36.92% Conservative: 30
Best Local Similarity: 22.90% Mismatches: 70
Query Match: 2.79% Indels: 65
DB: 13 Gaps: 12
US-09-802-285A-2 (1-659) x BJ411853 (1-663)
QY 269 GlyIleSerValLeuAsnThrGluIleLysLys-----GlnValTyrAla 283
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 96 GGTTCAGTTTAAATGACTACATCAATTAATCAGTTCCAATGGACCAACAATTTATAT 155
284 AspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspPheLeu 303
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 156 GAAGGTACAAATTTTTCATCAATGCCAGATATAAT-----191
304 LysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrVal 323
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 192 -----TATAATTTTGAATAAACCACCAACCCAGGAGTAATGTT 227
324 GlnThrValGluAsnMetIleMetAlaLeuIle-----SerIleSerLeuProAspTyr 341
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 228 AAAACCCCTT---GGAATGTTAAAGCAAAATTTGAATCAACCAACAATAATTCAGAACTT 284
342 AsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPhe 361
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 285 TCTTCA-----GATGATACTAATAACTTACCACAAATACATTAGTACTATGCAATAT 335
362 AlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAspGly 381
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 336 CCA-----GAGTTATTTAATAATATCACTTTACATCAAT 368
382 LysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeu-----394
369 GTTGATGCACCCAGCATCCAAAACCTCTGTTAAATAATTTACCAATGAAATACAAATCAA 428
395 -----SerAsnAlaGlyPheTyrThrPhe-----402
429 TTAGATCAACAGCAATAGTGGTTTATATCTTTTCAATCAAACTTTTCCCAATGAT 488
403 ArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProGlyGlu 422
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 489 AACATGGGCTACGATGTAGATCCATCATTTAGAAATTTATAAAACAAAGT-----GGTAAC 542
423 PheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLys---GlyArgAsnPhe 441
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 543 TATCAT-----AATTTCCATTTTGTGTAAAGATGAATAGTCGTTT 564
442 ThrProAspAlaGly---ValPheValTyrSerGlyAspGlu 454
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 585 ACTTATATGGAGGTGAAGTTTATTTTAAAGGTGATGAT 626

RESULT 9
BJ411853 663 bp mRNA linear EST 23-JUL-2002
LOCUS BQ752776
DEFINITION WHE4119_A05 B092S Wheat salt-stressed root cDNA library Triticum
aestivum cDNA clone WHE4119_A05_B09, mRNA sequence.
ACCESSION BQ752776
VERSION BQ752776.1 GI:21930558
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum

/strain="AX4"
/db_xref="taxon:44689"
/clone="adv6g05"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
/sex="mat A"
/dev_stage="Growth phase"
BASE COUNT 253 a -92 c 90 g 228 t
ORIGIN
Alignment Scores:
Pred. No.: 0.295 Length: 663
Score: 97.50 Matches: 49
Percent Similarity: 36.92% Conservative: 30
Best Local Similarity: 22.90% Mismatches: 70
Query Match: 2.79% Indels: 65
DB: 13 Gaps: 12
US-09-802-285A-2 (1-659) x BJ421576 (1-659)
/strain="AX4"
/db_xref="taxon:44689"
/clone="adv6g05"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
/sex="mat A"
/dev_stage="Growth phase"
BASE COUNT 253 a -92 c 90 g 228 t
ORIGIN
Alignment Scores:
Pred. No.: 0.295 Length: 663
Score: 97.50 Matches: 49
Percent Similarity: 36.92% Conservative: 30
Best Local Similarity: 22.90% Mismatches: 70
Query Match: 2.79% Indels: 65
DB: 13 Gaps: 12
US-09-802-285A-2 (1-659) x BJ411853 (1-663)
QY 269 GlyIleSerValLeuAsnThrGluIleLysLys-----GlnValTyrAla 283
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 96 GGTTCAGTTTAAATGACTACATCAATTAATCAGTTCCAATGGACCAACAATTTATAT 155
284 AspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspPheLeu 303
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 156 GAAGGTACAAATTTTTCATCAATGCCAGATATAAT-----191
304 LysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrVal 323
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 192 -----TATAATTTTGAATAAACCACCAACCCAGGAGTAATGTT 227
324 GlnThrValGluAsnMetIleMetAlaLeuIle-----SerIleSerLeuProAspTyr 341
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 228 AAAACCCCTT---GGAATGTTAAAGCAAAATTTGAATCAACCAACAATAATTCAGAACTT 284
342 AsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPhe 361
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 285 TCTTCA-----GATGATACTAATAACTTACCACAAATACATTAGTACTATGCAATAT 335
362 AlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAspGly 381
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 336 CCA-----GAGTTATTTAATAATATCACTTTACATCAAT 368
382 LysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeu-----394
369 GTTGATGCACCCAGCATCCAAAACCTCTGTTAAATAATTTACCAATGAAATACAAATCAA 428
395 -----SerAsnAlaGlyPheTyrThrPhe-----402
429 TTAGATCAACAGCAATAGTGGTTTATATCTTTTCAATCAAACTTTTCCCAATGAT 488
403 ArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProGlyGlu 422
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 489 AACATGGGCTACGATGTAGATCCATCATTTAGAAATTTATAAAACAAAGT-----GGTAAC 542
423 PheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLys---GlyArgAsnPhe 441
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 543 TATCAT-----AATTTCCATTTTGTGTAAAGATGAATAGTCGTTT 564
442 ThrProAspAlaGly---ValPheValTyrSerGlyAspGlu 454
|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 585 ACTTATATGGAGGTGAAGTTTATTTTAAAGGTGATGAT 626

RESULT 8
BJ411853 663 bp mRNA linear EST 10-MAR-2002
LOCUS BQ752776
DEFINITION Dictyostelium discoideum cDNA library, VF Dictyostelium
discoideum cDNA clone adv6g05.3, mRNA sequence.
ACCESSION BQ752776
VERSION BQ752776.1 GI:19324525
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 663)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLES Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshinigenes.nig.ac.jp.
Location/Qualifiers
1. .663
/organism="Dictyostelium discoideum"
FEATURES
source

```


REFERENCE
AUTHORS

3 Shibata, K., Itoh, M., Airawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed capillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

TITLE

JOURNAL
MEDLINE
PUBMEDREFERENCE
AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, D. M., Stauble, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P., Ring, B., Schöngard, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schombach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, J., Kohetsuki, S. and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851

TITLE

JOURNAL
MEDLINE
PUBMEDREFERENCE
AUTHORS

5 (bases 1 to 1050)
Adachi, J., Airawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, P., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

TITLE
JOURNAL

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGAGATCAAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAGCGCGCAATTAATTCGAGTTAATTAATTAATTCGCGCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3'

FEATURES
source

end: SstI. Host: SOLR.
Location/Qualifiers
1..1050
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="PANTOM DB:1700023E12"
/db_xref="MGD:MGI:1901580"
/db_xref="taxon:10090"
/clone="1700023E12"
/sex="male"
/tissue type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev stage="adult"
1..1050
/gene="Prss21"
53..943
/gene="Prss21"
/note="data source: MGD, source key: MGI:1916698,
evidence: ISS
protease, serine, 21
putative"
/codon_start=1
/protein_id="BAB24495.1"
/db_xref="GI:12839280"
/db_xref="MGD:MGI:1916698"
/translation="MALQSTYLOVDPEKPELQEPDLSGPCGHRTPSPVIGDDAEL
GRWPQSGIRVWGNHLCGATLLNRRVLTAAHCFQKNDPFDWTQVGFELTSPSLWN
LQAYNSRQIEDIFLSPKSYEQPNDAIKLSSPVYNNFQPICLNLSYKFNRT
DCWVTGWAIGDESLPQTVQVAIINNSCMNMYKKDFRTNMGDMVCACTP
EGKDACDGGSGGLACQDITVYQGVVSWGIGGGRHNRPGVTNISHYNNIQSTM
IRNGLRDPDPVPLLLFLTLA"
1032..1037
/gene="Prss21"
/note="putative"
1050
/gene="Prss21"
/note="putative"
/gene="Prss21"
/note="putative"

gene

CDS

polya_signal

polya_site

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best local Similarity:

Query Match:

DB:

US-09-802-285A-2 (1-659) x AK006271 (1-1050)

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB


```

QY 173 LysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTyrArgProLeuGluVal 192
Db 290 AAG-----GATAACGATCTCTTTCATCGG----- 313
QY 193 SerAspArgValGlnSerLeuProThrPheSerLeuPheValAsnSerProAlaPhe 212
Db 314 -----ACAGTCCAG-----TTTGTGTAGCTGCATCCAGGCCATCTCTC 352
QY 213 ThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnAlaAspTyrLeuSer 232
Db 333 TGGAAC-----CTACAGCCCTATTCCAACCGTTTACCAATAGAAATATTTTCTGAGC 406
QY 233 ThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAla 252
Db 407 CCCAAGTACTCGAGCAGTATCCCAATGACATA-----GCCCTGCTGAGCTGCATCT 460
QY 253 GlyValSerPheProGluPheLysAspSerProArgTyrArgGlnThrGlyIleSerVal 272
Db 461 CCAGTCACCTACATAAATCTTCATCCAGCCC-----ATCTGCCTC 499
QY 273 LeuAsnThrGluIleLysLysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerPro 292
Db 500 CTGAAGTCCACGTACAG-----TTTCAG----- 523
QY 293 IleTyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgVal 312
Db 524 -----AACCGAAGTACGCTGGTGGTGACCGGCTGGGGGCT-----ATT 562
QY 313 AsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAla 332
Db 563 GGAGAAAGTATGAGAGTCTGCCATCTCCC-----AACACTCTCCAGGAAGTCAGGTAGCT 616
QY 333 LeuIleSerIleSerLeu-----ProAspTyrAsnThrProMet 345
Db 617 ATTATCACACAGCATGTGTAAACCATATGTACAAAAAGCCAGACTTCGCGACGAACATC 676
QY 346 PheGlyAsp 348
Db 677 TGGGGAGAC 685

RESULT 11
BQ744898 671 bp mRNA linear EST 17-JUL-2002
LOCUS 946112B02.y1 946 - tassal primordium prepared by Schmidt lab Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION BQ744898
VERSION BQ744898.1 GI:21891685
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
TITLE Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
JOURNAL clade; Panicoideae; Andropogoneae; Zea.
COMMENT 1 (bases 1 to 671)
Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946112 row: B column: 02.
Location/Qualifiers
1..671
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassal primordium prepared by Schmidt
source
FEATURES

```

```

lab"
/tissue type="tassels"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab host="XL0LR"
/Note="Organ: tassels; Vector: HybridZAP; Site 1: EcoRI;
Site 2: XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybridZAP. Sample insert size range was 350 bp
to 3 kb with a 1 kb average."
BASE COUNT 181 a 169 c 165 g 156 t
ORIGIN

Alignment Scores:
Pred. No.: 0.62 Length: 671
Score: 95.00 Matches: 46
Percent Similarity: 36.04% Conservatve: 25
Best local Similarity: 23.35% Mismatches: 64
Query Match: 2.72% Indels: 62
DB: 14 Gaps: 9

US-09-802-285a-2 (1-659) x BQ744898 (1-671)
QY 7 LysArgIleIleValPheAlaValIleAlaLeuSerSerGlyAsnIleLeuAlaGlnSer 26
Db 116 AAGCGATTACAGATTTCTTATTGCTGCACATCTTCTGACAATGTCTG-----TAT 169
QY 27 SerSerIleThrArgLysAspPheAspHis-----IleAsnLeu 39
Db 170 ATCCAGGTCGGTGATTCCTGATCTGGACCAACAATTCTGGGAAAGGCCAGAAACAATGACC 229
QY 40 GluTyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAla 59
Db 230 GAGAAACACCTCTCACACAGATCAACAAGAGACTCCCGATCAGATGTTCTGCTGAG 289
QY 60 AlalysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSer 79
Db 290 GCAGCAGCAGCATCGGAGCTGCTTCAATGGTCTTCAAAATCCAGAGATACCACTATTCT 349
QY 80 AsnAlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAla 99
Db 349 ----- 349
QY 100 AspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGly 119
Db 350 GATGTGCTTCTTCAGCAT-----GCTCAGAAACTATTTCACCTTTTGCTGACACTTAT 400
QY 120 LysAspIleAsnTrpGlnMetTrpPro---ValLysAsp----- 131
Db 401 AGAGGCTCTCAAGCGAGAGCTATCCCAAGCTCCAGGATTTCTATATACTCCACCACTTAC 460
QY 132 ---AsnGluValArgTyrGlnLeuHisArgValLysTyrTrpGlnAlaMetAlaLeuVal 150
Db 461 GTTGATGAGCTTCTATG-----GCAGCGAGTTGGCTC 493
QY 151 TyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpValTyrGln----- 166
Db 494 TACCAGCGCCACAGCGCAGCAGCATCTATCTCAGCTATGTAACTGTACAGACGGGAAAGT 553
QY 167 TyrSerAspTrpAlaArg-----LysAsnPro 175
Db 554 TATGCTGATGGGAAGGCCAACATGTTGATGGGATGACAAAAATCCA 604

RESULT 12
CNS06JJZ 1001 bp DNA linear GSS 17-JUN-2001
LOCUS T7 end of clone AS0AA029D05 of library AS0AA from strain CLIB 533
DEFINITION of Saccharomyces bayanus, genomic survey sequence.
ACCESSION AL401685
VERSION AL401685.1 GI:12159698
KEYWORDS GSS.
SOURCE Saccharomyces bayanus.
ORGANISM Saccharomyces bayanus

```

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE AUTHORS

Souci t,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Boltin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Olier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvet,M.,
Wincker,P. and Weissenbach,J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

11152876

REFERENCE AUTHORS

Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
Aigle,M. and Durrens,P.

Genomic exploration of the hemiascomycetous yeasts: 5.

Saccharomycetes bayanus var. uvarum

FEBS Lett. 487 (1), 37-41 (2000)

20584715

11152880

REFERENCE AUTHORS

Direct Submission
Genoscope.

Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomycetes bayanus var. uvarum, Saccharomycetes
exiguus, Saccharomycetes servazii, Zygosaccharomycetes rouxii,
Saccharomycetes kluyveri, Kluyveromycetes thermotolerans, Kluyveromycetes
lactis var. lactis, Kluyveromycetes marxianus var. marxianus, Pichia
angusta, Debaryomyces Hansenii var. Hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES source

1. .1001
/organism="Saccharomycetes bayanus"
/strain="CLIB 533"
/variety="uvarum"
/db_xref="taxon:4931"
/clone="AS0AA029D05"
/clone_lib="AS0AA"
/note="end : 17"

<3. .>1001
/note="similar to Saccharomycetes cerevisiae ORF YNL262w [
POL2 : DNA-directed DNA polymerase epsilon, catalytic
subunit A]"

/evidence=not_experimental
350 a 161 c 209 g 280 t 1 others

BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 1.55 Length: 1001
Score: 94.00 Matches: 69
Percent Similarity: 35.45% Conservative: 48
Best Local Similarity: 20.91% Mismatches: 117
Query Match: 2.69% Indels: 96
DB: 17 Gaps: 19

US-09-802-285A-2 (1-659) x CNS06JUZ (1-1001)

QY 336 IleSerLeuProAspTyrAsnThrPrometPheGlyAspSerTirpileThrAsp----- 353
Db 108 TTGGATTCCCTGAATTAATATCCAGATATCATGATAATGGTACTAGTGTACT 167

QY 354 ---LysAsnPheArgMetAlaGlnPheAlaSerTirPalaArgValPheProLa----- 370
Db 168 GTTGATAATCTACACATAACACAAATATTGGCATCTGCATTCATCAATGATGCAGAGGT 227

QY 371 -----AsnGlnAlaIleLysTyrPheAlaThrAspGlyLysGlnGlyLysAla 386
Db 228 AOTGATCTCGTTAATAGCGGTATGGTATGGAGGATAAAATATCCACAGCACTCTACG 287
QY 387 ProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyrThrPheArg----- 403
Db 288 TCGGATTTTGTGAGGATGCGCTTTCTAATGATGCTTTAAATGCTCCTAAGGGCATGCTA 347
QY 404 ---SerGlyTyrAspLysAsnAlaThrValMetValLeuLysAlaSerProProGlyGlu 422
Db 348 AAGGAATGTTGGGATGAA----- 365
QY 423 PheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLysGly----- 438
Db 366 -----GCTTGAAGAGATAATGCACTGCGGATTTTGTGTAACACTCGTTAGCAAGCTGG 419
QY 439 ---ArgAsnPheThrProAspAlaGlyValPhe-----ValTyr 450
Db 420 GTCCAAAT-----CCTAATGCAAAATATTATTGATGTTTGTGAGATATCATGTCCAT 473
QY 451 Ser---GlyAspGluAlaIleMetLysLeuArgAsnTyrTyrArgGlnThrArgIleHis 469
Db 474 AACTTGACCAAAAAATCACTACTCCAATTAGTGAATGATTT-----AGCACTCTGGGC 527
QY 470 SerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTyr 489
Db 528 TCACTATCTGTATGTCAGATAGAACCAAAATTTTAAATCAAAACA-----ACCAAGTAC 581
QY 490 GluThrGlyAsnAsnLeuAspValLeuThrTyr-----ThrAsnPro 503
Db 582 TCGCCTGAAAATTTGTCAGCTTACAGTCAATATATCATGATAAGCGGTAGGACAAATCCA 641
QY 504 SerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLysTyrPheLeu 523
Db 642 ATGTTTCAGTTATTGGAT-----TTAAATATTAAAGCGTTATTGGGATCTG 686
QY 524 ValIle-----AspArg-----AlaIleGlyGluAlaThrGly----- 534
Db 687 TTAATATGATGATGATAAAATCAACTTCAGTGGTGTAGCTTTGTGGAAGATAGAGGAAAA 746
QY 535 -----AsnLeuGlyValHisTyrGlnLeuLysGluAspSerAsnProVal 549
Db 747 GAAAGCAAGACTATAACGCAATTTTCGGAATGGCACTAAGAAATTTTTCACCAATA 806
QY 550 PheAsp-----LysThrLys 554
Db 807 TATCAACCTGAATTTGAGGATTTGATGATGATCATCTTAGACAGTATGTTAAGACCAAG 866
QY 555 AsnArgValTyrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAla 574
Db 867 CAGAAC---TATTTCAATCAAAATTCAGGTACACAAAGCCTACCCAAATATCTCAACACG 923
QY 575 AspArgThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsnLysGluLeu 594
Db 924 AAGAAAGAA-----GACAAAGAGATGCG-----ATTGACAACTCACTG 962
QY 595 LysArgProAlaPheValPheGluLysPro 604
Db 963 AATGGGGTTCYACACATTTTCTCCAAGCCA 992
RESULT 13
BM413117
LOCUS
DEFINITION
EST587444 tomato breaker fruit Lycopersicon esculentum cDNA clone
CLEG62E22 5' end, mRNA sequence.
ACCESSION
BM413117
VERSION
BM413117.1 GI:18264747
KEYWORDS
EST.
SOURCE
tomato.
ORGANISM
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;


```
Query Match: 2.66% Indels: 60
DB: 13 Gaps: 11
US-09-802-285A-2 (1-659) x BML65623 (1-762)
QY 487 AsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrPro 506
Dd 17 AACAAAGAACACCAATAATGATCGATGACAGATTAATTAATGAGAAAAGTG 76
QY 507 AsnLeuAspHisGlnArgSerVal-----LeuPheIleAsnLys 519
Dd 77 AATGAATTACATCAAAATATATAGATTCTCAAGTGATCAAGCCCAATTAAGAAATAAG 136
QY 520 -----LysTyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGly 537
Dd 137 ATAACATAATATAGTACATATTGAT----- 163
QY 538 ValHisTrpGlnLeuLysGluAspSerAsn-----ProValPheAspLysThrLys 554
Dd 164 -----GATCTTAAAGAAAGAGATAATGAATTGTGGAAGTTATTAGCACATTAAAG 214
QY 555 AsnArgValTyrThrTyrArgAspGlyAsnLeu-----MetIle 569
Dd 215 GAGTATATCAACACATACAAAGGAAGACACAGATATTCCACCAATATGACGATCTAATA 274
QY 570 GlnSerLeuAsnAlaAspArgThrSerLeuAsnGluGluGluGlyLysValSer----- 587
Dd 275 AATAACTCTTAAAGAGATGTCACATTGAATGAAGAAATTCTTAAGCAAAAAAAG 334
QY 588 -----TyrValTyrAsnLysGluLeuLysArgProAlaPheValPheGlu 602
Dd 335 AATATAGTGTGATGTTACTATTACAAAGGAATTGCAA-----ATGTTGCAA 362
QY 603 LysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyrAsp--- 621
Dd 383 AATGATAAG-----ATGAAATATTATGAAAAATTTGATATA 418
QY 622 ---GlyGlnLysAlaPro-----GluIleSerIleArgGluAsnLys 634
Dd 419 GCAAATGATAGTGTCTCCCATTTATATGAGCCATTAAATGTAACATTACATGAAAAAAA 478
QY 635 GlyAsnAspPheGluLysGlyLysLeuAsn 644
Dd 479 GGAGAT-----GAAAAAATGAAAAAAT 502
RESULT 15
BI647560
LOCUS
DEFINITION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11807 row: m column: 10
High quality sequence stop: 654.
Location/Qualifiers
```

```
source 1. 780
/organism="Mus musculus"
/strain="129,C57BL/6J,FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5318577"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."
BASE COUNT 173 a 222 c 248 g 137 t
ORIGIN
Alignment Scores:
Pred. No.: 1.4 Length: 780
Score: 93.00 Matches: 30
Percent Similarity: 41.73% Conservative: 23
Best Local Similarity: 23.62% Mismatches: 34
Query Match: 2.66% Indels: 40
DB: 13 Gaps: 6
US-09-802-285A-2 (1-659) x BI647560 (1-780)
QY 135 ArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThr 154
Dd 261 AGGATGGAGATACACAGATTGCTCTCTGGACATCGATATCTGTG----- 305
QY 155 GlyAspGluLysTyrAlaArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsn 174
Dd 306 ---GGCCATCGATTCCCAAGATCATGGGCTTAGAGAGAGACAGGTTCCACGAGCGGCT 362
QY 175 ProLeuGly----- 177
Dd 363 CCGCTGGTCTCCAGTGTACGTGGATGACAACTGGGGGTGACTGTCAGTGGGCTTCTCTG 422
QY 178 LeuSerGlnAspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGln 197
Dd 423 CTTAGCAGCCAGATGATGCTGTCTCATCTGGAGGGACCAAGAAATGGTATGATCAAG 482
QY 198 SerLeuPro-----ThrPheSerLeuPheValAsnSer 209
Dd 483 CAGCTTCTCTCCGTGATGTAGACTGGGCGGAGACGTGGACCTAC---CTCATCGTAGGACAC 539
QY 210 ProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnGlnAlaAsp 229
Dd 540 CCGCTCTGGACAGT-----CTGACGAACACCTCTCAGTGGTCCGACG 581
QY 230 TyrLeuSer-----ThrHis 234
Dd 582 TACCTGGCTGGCGCACACAT 602
Search completed: August 4, 2003, 12:30:08
Job time : 2474 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 20:46:13 ; Search time 109 Seconds

(without alignments)
1560.153 Million cell updates/sec

Title: US-09-802-285A-2

Perfect score: 3494

Sequence: 1 MTTKFKRIIVFAVIALSSG.....KGLNLTITNGKQLVLVP 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3494	100.0	659	2 Q59289	Q59289 pedobacter
2	250.5	7.2	634	16 Q8E376	Q8E376 streptococc
3	250.5	7.2	634	16 Q8DXF4	Q8DXF4 streptococc
4	232	6.6	635	16 Q8K866	Q8K866 streptococc
5	231	6.6	635	16 Q9A0T7	Q9A0T7 streptococc
6	229	6.6	635	16 Q8P1V6	Q8P1V6 streptococc
7	191.5	5.5	645	2 Q8KN09	Q8KN09 streptomyce
8	189	5.4	633	16 Q8CZ74	Q8CZ74 streptococc
9	186.5	5.3	653	16 Q8RJK0	Q8RJK0 streptomyce
10	176	5.0	546	16 Q97SK5	Q97SK5 streptococc
11	176	5.0	614	2 Q8KN79	Q8KN79 pseudomonas
12	170.5	4.9	550	2 Q8KT36	Q8KT36 pseudomonas
13	160.5	4.6	612	16 Q8XR51	Q8XR51 raietonia s
14	160	4.6	1183	16 Q9GQH9	Q9GQH9 mycoplasma
15	154	4.4	713	2 O05937	O05937 pseudomonas
16	145.5	4.2	1165	16 Q8NZB0	Q8NZB0 streptococc

17	140.5	4.0	1165	16 Q8XSS2	Q8XSS2 streptococc
18	139	4.0	561	16 Q9X857	Q9X857 streptomyce
19	139	4.0	764	16 Q92A22	Q92A22 listeria in
20	137.5	3.9	604	2 Q9RMC9	Q9RMC9 acinetobact
21	137.5	3.9	1165	16 Q99XX8	Q99XX8 streptococc
22	133.5	3.8	1165	2 Q8KLP1	Q8KLP1 streptococc
23	132.5	3.8	1555	5 Q9U0N0	Q9U0N0 plasmodium
24	127	3.6	560	10 Q9CS12	Q9CS12 arabidopsis
25	126.5	3.6	696	16 Q8CP27	Q8CP27 staphylococ
26	126	3.6	1179	2 Q9AHK5	Q9AHK5 borrelia bu
27	125.5	3.6	672	16 Q8ZBE2	Q8ZBE2 yersinia pe
28	125	3.6	656	16 Q8CLP5	Q8CLP5 yersinia pe
29	125	3.6	1023	2 Q93T53	Q93T53 streptococc
30	124.5	3.6	1210	5 Q8IAV1	Q8IAV1 plasmodium
31	124.5	3.6	1883	16 Q9PPT2	Q9PPT2 ureaplasma
32	121.5	3.5	1088	16 Q8EUS7	Q8EUS7 mycoplasma
33	121	3.5	965	2 Q9AHL6	Q9AHL6 pasteurella
34	121	3.5	965	16 Q9CMP0	Q9CMP0 pasteurella
35	120.5	3.4	625	16 Q9ER34	Q9ER34 mycoplasma
36	120.5	3.4	992	16 Q65931	Q65931 mycobacteri
37	120	3.4	1046	2 Q84941	Q84941 streptococc
38	120	3.4	1272	16 Q98PR8	Q98PR8 mycoplasma
39	119	3.4	776	16 Q98PT2	Q98PT2 mycoplasma
40	118.5	3.4	579	3 Q9P6J3	Q9P6J3 schizosacch
41	118.5	3.4	839	16 Q8D518	Q8D518 vibrio vuln
42	118	3.4	772	2 Q46080	Q46080 pedobacter
43	117.5	3.4	675	5 Q81KA2	Q81KA2 plasmodium
44	117.5	3.4	693	5 Q94728	Q94728 riptortus c
45	117.5	3.4	2902	16 Q9ZME6	Q9ZME6 helicobacte

ALIGNMENTS

RESULT 1

Q59289 ID Q59289 PRELIMINARY; PRT; 659 AA.
AC Q59289;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Heparinase III protein precursor (EC 4.2.2.8).
GN HEPC.
OS Pedobacter heparinus (Flavobacterium heparinum).
OC Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales;
OC Sphingobacteriaceae; Pedobacter.
OX NCBI_TaxID=984;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96316388; PubMed=8702264;
RA Su H., Blain F., Musil R.A., Zimmermann J.J., Gu K., Bennett D.C.;
RT "Isolation and expression in Escherichia coli of hepB and hepC, genes
RT coding for the glycosaminoglycan-degrading enzymes heparinase II and
RT heparinase III, respectively, from Flavobacterium heparinum.";
RL Appl. Environ. Microbiol. 62:2723-2734(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Tkalec A.L.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U27586; AAB18278.1; --
KW Lyase; Signal.
FT SIGNAL 1 24 POTENTIAL.
SQ SEQUENCE 659 AA; 75806 MW; B73EDF10A1256FE2 CRC64;

Query Match 100.0%; Score 3494; DB 2; Length 659;
Best Local Similarity 100.0%; Pred. No. 9.1e-232;
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTTKFKRIIVFAVIALSSGNILAQSSITRKDFDHINLEYSGLEKVKVKAAGNYDDAA 60

Db 1 MTTKFKRIIVFAVIALSSGNILAQSSITRKDFDHINLEYSGLEKVKVKAAGNYDDAA 60

QY 61 KALLAYREKSKAREPDFSNAEKPADIRQPIDKVTREMAADKALVHQFQPHKGYGIFDYK 120

```
Db 61 KALLAYREKSKAREPDSNAEKPADIRQPDVKVTREMDKALVHQFQPHKGYGFDYCK 120
QY 121 DINQWMPVKDNEVWQHLVRVWQAMALVYHATGDEKYAREWVYQYSDWARKNPLGLSQ 180
Db 121 DINQWMPVKDNEVWQHLVRVWQAMALVYHATGDEKYAREWVYQYSDWARKNPLGLSQ 180
QY 181 DNDKFWPRLVSDRVOSLPPTFSLFVNSPAPTPAPLMEFLNSYHQADYLSLTHVAEQN 240
Db 181 DNDKFWPRLVSDRVOSLPPTFSLFVNSPAPTPAPLMEFLNSYHQADYLSLTHVAEQN 240
QY 241 HRLFEAQRNLFAVGFPEFKDSPRWQRTGISVINTKQVYADGMQFELSFIYHVAID 300
Db 241 HRLFEAQRNLFAVGFPEFKDSPRWQRTGISVINTKQVYADGMQFELSFIYHVAID 300
QY 301 IFLKAYGAKRVNLKEFPQSVQTVENMINIMALISISLPDYNTPMGDSWITDKNFRMAQ 360
Db 301 IFLKAYGAKRVNLKEFPQSVQTVENMINIMALISISLPDYNTPMGDSWITDKNFRMAQ 360
QY 361 PASWARPVPANAIKYFATDGKQKAPNPLSKALNAGYTFERSGWDKNATVMVLKASPP 420
Db 361 PASWARPVPANAIKYFATDGKQKAPNPLSKALNAGYTFERSGWDKNATVMVLKASPP 420
QY 421 GEFHAQPDNGTPELFPIKGNFTPDAGVFVYSGDEAIMKLRNMYRQTRIHSITLTDNQNV 480
Db 421 GEFHAQPDNGTPELFPIKGNFTPDAGVFVYSGDEAIMKLRNMYRQTRIHSITLTDNQNV 480
QY 481 ITKARQNKWETGNNLDVLTYPSPNLDHORSVLPINKKYFLVDRAIGBATNGLGVHW 540
Db 481 ITKARQNKWETGNNLDVLTYPSPNLDHORSVLPINKKYFLVDRAIGBATNGLGVHW 540
QY 541 QLKEDSNPVDKTKRNVYTYRDGNNLMIOSLNADRTSLNEEGKVSYYNKKELKRPAPV 600
Db 541 QLKEDSNPVDKTKRNVYTYRDGNNLMIOSLNADRTSLNEEGKVSYYNKKELKRPAPV 600
QY 601 FEKPKKNAGTQNFVSIVPYDQKAPETISIRENKGNDPEKGLNLTTLTINGKQQLVLP 659
Db 601 FEKPKKNAGTQNFVSIVPYDQKAPETISIRENKGNDPEKGLNLTTLTINGKQQLVLP 659

RESULT 2
QBE376 PRELIMINARY; PRT; 634 AA.
AC QBE376;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN GBS1885.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MEM316 / Serotype III;
RX MEDLINE=2242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst S.
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease."
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766854; CAD47544.1; -
DR Sagalists; gbs1885; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 634 AA; 74562 MW; A31C360CDDC0B711D CRC64;

Query Match
Best Local Similarity 22.8%; Pred. No. 8.1e-09;
Matches 127; Conservative 74; Mismatches 223; Indels 133; Gaps 21;
QY 123 NWQM-----W---PVKDN-EVRQHLHVRVWQAMALVYHATGDEKYAREWVYQ 166
```

RESULT 3

```
QBDXF4
ID QBDXF4 PRELIMINARY; PRT; 634 AA.
AC QBDXF4;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN SAG1897.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Kessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae."
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014277; AAN00759.1; -
DR TIGR; SAG1897; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 634 AA; 74622 MW; BFF2C40CDDC1E84FD CRC64;
```



```
QY 239 GNERLEAQRNLFGAGYSEPEFKDPRWQRTGISVLNTEIKKQVYADGMQFELSPIYHV-- 236
Db 181 SNWGIILQTPMLAIYHFFSDKMDLEAYHPASBELKQOIETQILGDSQFEQSIYHVEV 240
QY 297 --AAIDIFLKAYGSAKRVNLEKFFQSYVOTVENMIMALISISLPDYNTPMFGDSWITDK 354
Db 241 YKALLDLCL-----LLPDLQDSYQELLEKMKATYIQMTGLDORTLAFGDSSTET 230
QY 355 NFRMAQFA-----SWARVFPANAIKYPATGKQKQKAPNFKLSKALSNA 337
Db 291 TEILSLSAVLVNQEIDLNGLDVVDLSLLFLGREKVKRLQEFKQWQPK--SMIFEDS 348
QY 398 GFYTFRSQWGNKATVMVLKASPGEPHAQDNGTFFELFKGRNFTPDAGVFVYSGDEAIM 457
Db 349 GHVCIKD---EHRYLFFKNGPLGSAHSHSDENSFCLOYGQPIFIDAGRYSY----- 397
QY 458 KLRNWRQTRI-----HSTLTLDNQNMVITKARQ---NKWE-----TCGNLDVLT 499
Db 398 --REIYERYLLKSAWHSSTCIVDG-----KAPERITGSWEYEPYPSLFCCHKEREQVH 449
QY 500 YNPSY-----PNLD--HORSVLFIKXYFLVID--RAIGBATGNLGVHMKEDSNPVED 551
Db 450 YIEGAYWSAEPDLPYLHRKILMLVEDVWLLVDIDRCQGO-----HEVLTOF---ILD 499
QY 552 KTKNRVYTYRDG--NNLMIQS 571
Db 500 KD-----VTYQDKINQLRLWS 516

RESULT 11
Q8KN79
ID Q8KN79 PRELIMINARY; PRT; 614 AA.
AC Q8KN79
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF 12.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22053227; PubMed=12057956;
RA Raymond C.K., Sims E.H., Kas A., Spencer D.H., Kutyavin T.V.,
RA Ivey R.G., Zhou Y., Kaul R., Clendenning J.B., Olson M.V.;
RT "Genetic Variation at the O-Antigen Biosynthetic Locus in Pseudomonas
RT aeruginosa.";
RL J. Bacteriol. 184:3614-3622(2002).
DR EMBL; AF498414; AAM27770.1; -.
DR EMBL; AF498414; AAM27770.1; -.
KW Transferase.
SQ SEQUENCE 614 AA, 69144 MW, 2E44A45D87511F00 CRC64;

Query Match 5.0%; Score 176; DB 2; Length 614;
Best Local Similarity 23.0%; Pred. No. 0.001;
Matches 100; Conservative 55; Mismatches 178; Indels 102; Gaps 24;

QY 113 YGYEDY-----GKIDNQWQWPKNEVR-----WQLHRVQWQA 146
Db 72 FGYPVYADSVDPHNSNLTGVMTNSDRNW--WEIPDFDLAVGDIKAYWEASRFDW--V 127
QY 147 MALVYHATGDEKYARE---WVYQYSDWARKNPLGLSQDNDFVWR--PLEVSDRVQSLEPP 201
Db 128 PALAQKALAGESGALDKINAWL---NDWCNPNPPYKGN-----WKCGQEASIRVMHIAL 179
QY 202 TFSLF--VNSPAPTFALMEFLNSYHQADYLSYTHY--AEQGNHRLFEAQRNLFAGVSFP 257
Db 180 AAILLCQVTPA---PAALLDLAR--IHLRIVPTIRVAIAQDNHNGTSEAAA--LFIGGSWL 234
QY 258 EFKDSP---RWQRTGISVLNTEIKKQVYADGMQFELSPIYHVAIDIFLKAYGSAKRVNL 314
Db 235 RAHGEPDAVTWMLRGLKLENRAKHLIGEDGFSQVSLNHYRVMVLDTFSLAEIWRRLQLSL 294
```

```
QY 315 EKEFP---QSYVOTVENMIMALISISLPDYNTPMFGDS-----WITDKNFR-- 357
Db 295 -VEFSARWYSRASAATHWLHAFID-----PATGDGPNLGDGARGLLPLTQTDYRDY 345
QY 358 -----MAQFASWARVFPANQ---AIKYFATDGKQKAPNFKLSKALSNAAGFYTFPSGW 406
Db 346 RPAVOLAMALFTD--TWAYFGNCECSPILLGLDLFSAESPVPRSRVDFDEGGTAILRSG-- 403
QY 407 DKNATVMVLKASPGGFFH--AQPDNGTFFELFKGRNFTPDAGVFVYSGDEAIMKLRNWRQ 465
Db 404 ----QAMAMLRVPRFRPRFSQADALHLDLWLEGRNLLRDAGTYSYNTESAWL---SVFPG 456
QY 466 TRIHSTLTLDNQNMV 480
Db 457 TASHNTIQFDGRDQM 471

RESULT 12
Q8KI36
ID Q8KI36 PRELIMINARY; PRT; 550 AA.
AC Q8KI36
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Similar to prenyltransferase and squalene oxidase.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22053227; PubMed=12057956;
RA Raymond C.K., Sims E.H., Kas A., Spencer D.H., Kutyavin T.V.,
RA Ivey R.G., Zhou Y., Kaul R., Clendenning J.B., Olson M.V.;
RT "Genetic Variation at the O-Antigen Biosynthetic Locus in Pseudomonas
RT aeruginosa.";
RL J. Bacteriol. 184:3614-3622(2002).
DR EMBL; AF498401; AAM27563.1; -.
DR EMBL; AF498411; AAM27714.1; -.
KW Transferase.
SQ SEQUENCE 550 AA, 61874 MW, 355BAE0565FDDDB6D CRC64;

Query Match 4.9%; Score 170.5; DB 2; Length 550;
Best Local Similarity 24.0%; Pred. No. 0.0021;
Matches 88; Conservative 46; Mismatches 151; Indels 81; Gaps 18;

QY 144 WQAMALVYHATGDEKYAREWVY--QYSDWARKNPLGLSQDN--DKFV----- 186
Db 80 WQA-----DRSSKLWLYNLVLDLNLNARDIG--SOPGLADKLIQSWIQANPPVSGE 128
QY 187 -WRPLEVSDRVOSLPPTFSLFVNSPAPTFALMEFLNSYHQADYL--STHYAEQGNHRL 243
Db 129 GWEPYPLSLRIVNLVKWLARHDERSTF-----LADSLAVQADALVQOVVEYHILGNHLF 181
QY 244 FEAQRNLFAGVSFPFKDQSPRWQRTGISVLNTEIKKQVYADGMQFELSPIYHVAIDIFL 303
Db 182 ANGKALVFAAG--AYLSGAMADRWLAKRLDLDELPEQFLNDGGHPELSMPYHATLL----- 236
QY 304 KAYGSAKRVNLE-----KEFFQSYVOTVENMIMALISISLPDYNTPMFGDSWITDK 354
Db 237 --WDMCDLVNLSRGLPDLAERLPQ--WREVVVQGLKWLRSQHDPDGRISFND----- 287
QY 355 NFRMAQFASWARVFPANAIKYFA--TDGKQKAPNFKLSKALSNA--GFYTFRSQWGNKAT 411
Db 288 ----AAFG-----IAPEYEDIAAYAKRLDISPPAHENHLLAAIYNSATGVAALPADGVKAI 339
QY 412 VMVLKASP---PGEFHAQPDNGTFFELFKGRNFTPDAGVFVYSGDEAIMKLRNWRQTRI 468
Db 340 LDLAKVGPDYQPG--HAHADTLSPFELSVFGRRLVYNSGTSTQYGGDSE----RQRQGTAA 393
QY 469 HSTLTLT 474
Db 394 HNTVGL 399
```

Db 538 SAVLRWELCPGAQAGNVVQSEAGNLVTADVPIRRIELTSGRESRYYLQTEPIVPLEV- 598

Qy 591 NKEIKRP 597

Db 597 --ELDRP 601

RESULT 14

Q98QH9 PRELIMINARY; PRT; 1183 AA.

AC Q98QH9;

DT 01-OCT-2001 (TrEMBLrel. 18, Created)

DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)

DE Hypothetical protein MYPU_3820.

GN MYPU_3820.

OS Mycoplasma pulmonis.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=2107;

RN [1]

SEQUENCE FROM N.A.

RC STRAIN=UAB CTIP;

RX MEDLINE=21267165; PubMed=11353084;

RA Chambaud I., Hellig R., Ferris S., Barbe V., Samson D., Galissou F.,

RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,

RA Blanchard A.,

RT "The complete genome sequence of the murine respiratory pathogen

RT Mycoplasma pulmonis.";

RL Nucleic Acids Res. 29:2145-2153 (2001).

DR EMBL; AL445564; CAC13555.1; -

KW MypuList; MYPU_3820; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 1183 AA; 140128 MW; 565C3A6949E1A9D0 CRC64;

Query Match 4.6%; Score 160; DB 16; Length 1183;

Best Local Similarity 18.7%; Pred. No. 0.033;

Matches 159; Conservative 131; Mismatches 252; Indels 308; Gaps 44

Qy 7 KRIT--VEAVIAL-----SSGNILAQSSSITR-----KDFD----- 35

Db 4 KIIILASFLFMSLFLFSSNSTIAQKNITQVFKNSLENDFYRKNNDEIGEKTLS 63

Qy 36 -----HINLBSGLKVNKVAAGNYDDAAKALLAYIREKSKAREPDFSNAEKPAD 86

Db 64 IVSKFIETKNLNIYKDFEKINYIVNDSNE-----NFSLGNNSQ 105

Qy 87 IRQPIDKVTREMAKALVHQPHQPHKGYGFDYGDKNQWMPVKDNEVRWLHVRUKWQA 146

Db 106 IE-----NNTNLNVKNTDID-KLNSVKFF-- 128

Qy 147 MALVYHATGDEKAREWYQYSDWARKNPLGLSQNDKDFVWRPLSVSRVQSLPPTFSLF 206

Db 129 -----FNDKSSQSW-----DFIEITWIKQRIEKISDLSQK 160

Qy 207 VNSPAFTA-----FLME-----PLNSYHQADYL-----STHYAQGNHRLFEAQ 247

Db 161 YNDNHFSVSIKKLDFYIDYKSPKNDSLFNEFRKKVDVLEIKRNVFVYKEINSQTKIV 220

Qy 248 RNLFAGVSFFPEFKDSPR--WRQ-TGISLVNTEIKKQVYA--DGMQFELSPYHVAADI 301

Db 221 KD-FDKLNFELFTSSLLFWKENVNLDLHNEIYKIQALDPENKKNFEIK--YQILSDDL 277

Qy 302 FLKAYGSAKRVNLBKEFPQSY--VQTVENMTMALISISLPDNT----- 343

Db 278 I-----SSDVKFLLLFRGEYFNQNIIDLGIKRVKISRDSINSQPIKQRLLIYVKNV 331

Qy 344 -----PMF-----GDSWITDKNFERMAQFASWAR---VEPNAQAIK 375

Db 332 YKNFDGKNGEKEVPQFIHEEREITSNEYGGHVLNQQFFRMAFISDONNEVIYVDDIAD 391

Qy 376 -----YFATDGKQCKAPN-----FLSKALSNAGFYTRSGWD--KNA 410

Query Match	4.6%	Score 160.5	DB 16	Length 612
Best Local Similarity	21.3%	Pred. No. 0.012	Indels 119	Gaps 31
Matches 129	Conservative 85	Mismatches 274		
QY 49	KAVAAGNYDDAAKALLA-----YYREKSKAREPDF-SNAEKPADIRQIDKVTREMDAKA	102		
DB 56	KAAPEGLVGHAVEETVAVRVFGWYQPAQA-EPDHRNPFNGARVNEP-SQPWWQIAD--	111		
QY 103	LVIHQFQPHKGYPDYCKDINWQMPYKDNVRLHVKRWQWQAMALVYHATGDEYA--	160		
DB 112	-----FDPVAG-----DIKAIWEASRFDW--VLIVLAQQAVRGEPQAMT	147		
QY 161	--REWVYOYGDWARKNPLGLSQDNDKFWVR-PLFVSDRVQSLPPTFSLFVNSPAFTPAFL	217		
DB 148	QLANWL---TSWARANPPYLGN-----WKCGQEAALRVHMLAALVLGHFSAAPA-L	198		
QY 218	MEFLNSVHQQ-ADYLSLTHYABQGNHRLFEAQRNLPAGVSPFEPKDSF--RWRGTGIVSL	273		
DB 199	MALVVRTHLQRIAPTGLVAYADQNNHGTSEAAA-LFIGGSLAAQCPDGRHWHQAGSHWL	257		
QY 274	NTEIKKQVADGMQFELSPYVHVAIDIFLKAYGSAKVNLEKEFP-----QSVQVQIVE	327		
DB 258	ENRARKLIADGSPFSQHSVMYH-----RLMLDTYSMAEVRWRHWSLPAFSAQLQARLGGAS	313		
QY 328	NMIVALISILPDYNTPMFGDS-----WITDKN-----FRMAQFASKARV	367		
DB 314	NWLYQMIDRTTGD--APNLGANDGARLLPLTADHRDFRPSVLQALCALFORADAFGWDGE	371		
QY 368	FPAQAIAKYATDCQKQKAPNFLSKALSNAAGFYTFRSGWDKNATVWLKASPPGEFH--A	425		
DB 372	W--SDAURLVGVPRPEQVRPPARSHMEAGYGILLRYG---RAPALF---NLPHRRHPS	423		
QY 426	QPQNGTPELETIKGRNFTPDGAVFYVSGDEATMKLNRNVRQTRISTLTLDQNQNV--ITK	483		
DB 424	QADALHVDVFWLGGKNLLRDAGSFYSAAES---AGVFSGTASHNTVQFDQDQMPRLSR	479		
QY 484	ARQKWETGNLVLVITYNP-----SYPNLDHQBSVLFINKKFLVLDRAIGEATG	534		
DB 480	FLFGAWLTKARDVEPVKQTADGVTCAAGYVDQOGASHIRA-LTLGKRSLRVYDR-VGGERR	537		
QY 535	NLGVHWOL----KEDSNPVFOKTKNRVVTYTRDGNLMIQSLNADRTSLNEEGKVSVY	590		

Db 392 NIDNLFDFLEHQQRKKTKITIKTKDNTKNTLYEIIYSKELKDSALKARWENWDPINNL 451
QY 411 TVMVL-----KASPPGFHQAQDNGTGFELFI-----KGRNFTPD-----AGVF 448
Db 452 EHKLVKDNKNGKVPKNPRINPHNGLEBKIFWLEHEKLNHLSNFKKVPFNKEGAF 511
QY 449 V-----YSGD-----EAIMKLRNVYRQRIHSTLTLDNONNVITKARONKW-ETGNL 495
Db 512 VKYSISKAYSIDLPPENSLTLKKYFWNL-----SNDFSQEELITIKDNNSFSKSGNFL 568
QY 496 DVITYN-----PSYNLDHQRSLVLFINKYFLVIDRAIGEAATGNLGVHWOLK--- 543
Db 569 LELKNTNYFNSYNLISVGSYNLKHNL--IDKKILVPVEKSI---AGNILKNYLLKNFN 623
QY 544 -EDSNPFVDTKRNVYTYVDGNNLMQISNAARTSLNEZ-----EGKV-----SYVYNKE- 593
Db 624 FEES--ISSLSYBEIVQNKKLNVNLFIEFLNKNKYNKKIFDIENKMDSLSYVDNKNK 681
QY 594 ----LKRPAFVFEKPKKNAGTQNFVSIYPYDQKAPETISIRENKGNDPEKGLN---LT 646
Db 682 LISLDEEFKINSFKN--TKNFVEIDYQD---HPNYDLTFEKSIFYKDKINKKQIT 736
QY 647 LITNGKQQLV 656
Db 737 ININ-KERII 745

RESULT 15
O05937 PRELIMINARY; PRT; 713 AA.
AC O05937;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Algininate lyase.
GN ALYLL.
OS Pseudomonas sp. (strain OS-ALG-9).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=86038;
RN [1]__
RP SEQUENCE FROM N.A.
RC STRAIN=OS-ALG-9;
RA Kraiwattapanong J., Ooi T., Kinoshita S.;
RT "Nucleotide sequence and expression of a gene (alyII) for an algininate
lyase from Pseudomonas sp. OS-ALG-9."
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003330; BAA19848.1; -.
KW Lyase.
SQ SEQUENCE 713 AA; 79805 MW; 2CE6CD2895899F2E CRC64;

Query Match 4.4%; Score 154; DB 2; Length 713;
Best Local Similarity 20.3%; Pred.No.0.041;
Matches 153; Conservative 92; Mismatches 234; Indels 276; Gaps 42;
QY 87 IRQPIDKVTREMADKALVHQF---QPHKGYGDFYGDYKINWQWVPVKDNEVRWQLHRVKW 143
Db 48 LRQSYQAV-KNAADKALAQPIVVPVPEKGGGVTH-----EQHKNY 88
QY 144 WQAM-ALVYHATGDEKYA-----REWYQYSDWARKNPLGLSDN-----DKF 185
Db 89 SNMLNGVAYQISGEKKYADYVKNVWMLNAYASQYQWKP-LHPKRKSEEDGGRIFWQSLNDF 147
QY 186 VWR--PLEYSDRV-QSLPPTFSLFVNSPAPTAFLMEF-----LNSYHQQADYLSTH 234
Db 148 VWQLYTIQAYDLVYDIPADTKRTIEKLFVP--ILKFFTEDRYDVFNKIH----- 196
QY 235 YAEQGNHRLFEAQRNLFA-GVS-----FPE--FKDSPRWRQTGISVLNTEIKKQVYA 283
Db 197 -----NHGTW-----NLAAVGITCYVLNKEEYVEMAIKSGKDKGTGL---AQIDQLFSP 244
QY 284 DGMQFELSIYHVAIDIFPKAYGSAKRVNLEKEPPQSYVQTVENMIMALISISLPDYN 343
Db 245 DGYME-GPYQRYALLPFVL---FAKAIN-NVEPSRKIFEYRDKLLSKAIHTSL----- 294

QY 344 PMEGDSWITDKNFRMAQAFASWARVFPANQAIK----- 375
Db 295 ----QTSYTDKTP-----FPLNDAIKDKTYESVELVYGVDLAYADIKAEVDLLD 339
QY 376 -----YFATDGKQKAPNELSKAL-----SNAGFYTFERSGWDKNATV 412
Db 340 IARQQNRVIVSDAGLKVAADLAAGKAVPKYQTLWIRDDGKGDEGGGLRNGENTDQOC 399
QY 413 MVLKASPPGGEFHAQPD-----NGTFELFIKGRNFTPDAG-----VFVYSGDEAIMKL 459
Db 400 VVLKAASOGMGHGHFDRLNLLFYDNTTEIF-----PDYGAARFLNIDTKNGGGVLPEN 452
QY 460 RNWYRQTRIHSITLTD-----NQNMVITKARONKMET----- 491
Db 453 NTWAKQTVAHNALVVDQTSHFNAKLGPADKASPTLLYFNSQPNLKVVSADKAYTDVTM 512
QY 492 -----GNLADVLTNTNPSYNLDHQRSLVLF-----INNKYFLVID-----BAIGE 531
Db 513 LRTSALVKVEGLDKPLIIDVMQASAKSHQYDLPFWYKGOLVNTSPPTAKANQLTALGD 572
QY 532 ATGNLGVHWQLKEDSNPFVDKT-----KNRVYTTYTRDGN--LMIQSLNADRTSLNE 581
Db 573 KNGYQHI-W--LNASNPLEGKSGMVGLLNKNRYTTHFVSDNPLEVKLLSIGANDEPMNL 629
QY 582 BEGKUSYVYNKELKRPAPFVEKPKKNAGTQNFVSIYPYDQ-----OKAPEIS--- 629
Db 630 VDGK-----AFMLSSSQN---QTFVSIETHGGTDPINETVSSALPTVSGLK 674
QY 630 -IRENKGNDFEKGLN--LTLTIN--GKQQLVLV 658
Db 675 LIKSDAQQTIIISFKVNERITYTYQINTEKQQLYII 709

Search completed: July 28, 2003, 20:52:10
Job time : 114 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 20:31:37 ; Search time 26 Seconds
(without alignments)
1191.946 Million cell updates/sec

Title: US-09-802-285A-2

Perfect score: 3494

Sequence:

1 MTTKIFKRIITFAVIALSSG.....KGKLNLTITNGKQLVLP 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123	3.5	532	1 INV2 YEAST	P00724 saccharomyc
2	120	3.4	677	1 SACC_BACSU	P05656 bacillus su
3	119	3.4	1138	1 C7AB BACUA	Q45707 bacillus th
4	117.5	3.4	995	1 DPO1 KLULA	P09804 kluyveromyc
5	116	3.3	4563	1 APB HUMAN	P04114 homo sapien
6	115.5	3.3	1113	1 Y14D MYCPN	P75033 mycoplasma
7	115.5	3.3	1314	1 SS22 YEAST	P25390 saccharomyc
8	115	3.3	1063	1 HGPI HAEIN	P44795 haemophilus
9	114.5	3.3	1167	1 CLAA BACTU	P56956 bacillus th
10	114.5	3.3	1211	1 EX5B HAEIN	P45157 haemophilus
11	113.5	3.2	1398	1 PLS PYRFU	P72186 pyrococcus
12	113	3.2	2179	1 POLG EC23W	Q73556 e genome po
13	112	3.2	532	1 INV1 YEAST	P10594 saccharomyc
14	112	3.2	532	1 INV4 YEAST	P10596 saccharomyc
15	112	3.2	1138	1 C7AB BACUK	Q45708 bacillus th
16	111.5	3.2	1002	1 TAGA VIBCH	P24019 vibrio chol
17	110.5	3.2	473	1 SACC BACST	P94468 bacillus st
18	110.5	3.2	850	1 PRSN PIERA	Q9U8q4 pieris rapa
19	109.5	3.1	965	1 AMPN RABIT	P15541 oryctolagus
20	109.5	3.1	2334	1 WAPA_BACSU	Q07833 bacillus su
21	109	3.1	951	1 HEX ADE05	P04133 human adeno
22	106.5	3.0	819	1 CSDI ECOLI	P53512 escherichia
23	106.5	3.0	850	1 PRSN PEBR	Q9GV36 pieris bras
24	106.5	3.0	1104	1 COLA_CLOPE	P43153 clostridium
25	106	3.0	1132	1 TERT OXYTR	O76332 oxytricha t
26	106	3.0	1455	1 GTFC STRMU	P13470 streptococ
27	105	3.0	473	1 SACC_BACSU	P05655 bacillus su
28	105	3.0	529	1 UDBE HUMAN	O75310 homo sapien
29	105	3.0	1042	1 T1RH METJA	O60295 methanococ
30	105	3.0	1180	1 PC12 MOUSE	O55134 mus musculu
31	105	3.0	1271	1 Y338 MYCGE	P47580 mycoplasma
32	104.5	3.0	1207	1 DPOL ASF7B	P42489 african swi
33	104.5	3.0	1750	1 Y832_METJA	Q58242 methanococ

34	104	3.0	537	1 TEE6_STRPY	P18481 streptococc
35	104	3.0	802	1 CSD2_ECOLI	P53513 escherichia
36	103.5	3.0	543	1 APPA_BACSU	P42061 bacillus su
37	103.5	3.0	1183	1 CNA STAAU	Q53654 staphylococ
38	103.5	3.0	1433	1 SUBE_BACSU	P16397 bacillus su
39	103	2.9	674	1 ETE1_MSEPV	Q9YVX9 melanoplus
40	103	2.9	767	1 TOP1_MOUSE	Q04750 mus musculu
41	103	2.9	837	1 HFC1_HAEIN	P33397 haemophilus
42	103	2.9	917	1 SYI STAAU	P41972 staphylococ
43	103	2.9	1034	1 BGAL_BACME	O52847 bacillus me
44	102.5	2.9	2136	1 YCF2_MARPO	P09975 marchantia
45	102	2.9	513	1 YNM6_YEAST	Q03104 saccharomyc

ALIGNMENTS

RESULT 1
INV2 YEAST STANDARD; PRT; 532 AA.
AC P00724;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Invertase 2 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 2)
DE (Saccharase).
GN SUC2 OR YIL162W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
CX NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=83168934; PubMed=6300785;
RA Taussig R., Carlson M.;
RT "Nucleotide sequence of the yeast SUC2 gene for invertase.";
RL Nucleic Acids Res. 11:1943-1954(1983).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX PubMed=9169870;
RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
Moule S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
Skellton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
RL Nature 387:84-87(1997).
[3]
RN SEQUENCE OF 1-30 FROM N.A.
RX MEDLINE=85137467; PubMed=6396505;
RA Sarokin L., Carlson M.;
RT "Upstream region required for regulated expression of the glucose-repressible SUC2 gene of Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 4:2750-2757(1984).
[4]
RN SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=87064535; PubMed=3537718;
RA Kaiser C.A., Botstein D.;
RT "Secretion-defective mutations in the signal sequence for Saccharomyces cerevisiae invertase.";
RL Mol. Cell. Biol. 6:2382-2391(1986).
[5]
RN SEQUENCE OF 21-532, AND CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=88213364; PubMed=3284881;
RA Reddy V.A., Johnson R.S., Biemann K., Williams R.S., Ziegler F.D.,
Trimble R.B., Maley F.;
RT "Characterization of the glycosylation sites in yeast external invertase. I. N-linked oligosaccharide content of the individual sequons.";
RL J. Biol. Chem. 263:6978-6985(1988).
[6]
RN ACTIVE SITE ASP-42.
RP

```

RX MEDLINE=90293006; PubMed=2113524;
RA Reddy V.A., Maley F.;
RT "Identification of an active-site residue in yeast invertase by
RT affinity labeling and site-directed mutagenesis.";
RL J. Biol. Chem. 265:10817-10820(1990).
CC - CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC - PFM: THE ENZYME EXISTS IN TWO STATES: A GLYCOSYLATED EXTERNAL
CC FORM, OR A NON GLYCOSYLATED INTERNAL FORM.
CC - MISCELLANEOUS: THE SEQUENCE OF THE SECRETED FORM OF THE ENZYME IS
CC SHOWN.
CC - SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z46921; CAAB7030.1; -
DR EMBL; V01311; CA24618.1; -
DR EMBL; K03294; AAA35127.1; -
DR EMBL; M13627; AAA35129.1; -
DR EIR; A00899; IFBY.
DR GlycoSuiteDB; P00724; -
DR COMPLEYEST-2DPAGE; P00724; -
DR SGD; S0001424; SUC2.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005576; C:extracellular; IDA.
DR GO; GO:0004564; F:beta-fructofuranosidase activity; IMP.
DR InterPro; IPR001362; Glyco_hydro_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR SMART; SM00640; Glyco_32; 1.
DR PROSITE; P500609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Multigene family; Signal.
FT SIGNAL 1 19
FT CHAIN 20 532
FT CHAIN 21 532
FT ACT_SITE 42 42
FT CARBOHYD 23 23
FT CARBOHYD 64 64
FT CARBOHYD 97 97
FT CARBOHYD 111 111
FT CARBOHYD 118 118
FT CARBOHYD 165 165
FT CARBOHYD 266 266
FT CARBOHYD 275 275
FT CARBOHYD 356 356
FT CARBOHYD 369 369
FT CARBOHYD 384 384
FT CARBOHYD 398 398
FT CARBOHYD 512 512
FT MUTAGEN 42 42
FT CONFLICT 409 409
FT SEQUENCE 532 AA; 60639 MW; 7DBAB3B6772B775 CRC64;
Query Match 3.5%; Score 123; DB 1; Length 532;
Best Local Similarity 18.9%; Pred. No. 0.46;
Matches 109; Conservative 76; Mismatches 179; Indels 214; Gaps 31;
QY 94 VTRMADKALVHOFQPHKGV-----GYFDYKGDINWQW-PVKDNEVRWQLHRVKWQAM 147
Db 21 MTNETSDRLVH-FTPNKGMNDNGLWYDEKAKHLYFYQNPNTVW--GTFPLWG-- 75
QY 148 ALVYHATGDEKYAREWYQYSDWARKNPLGLSQNDKFKVWRPLEVSDRVQSLPTTSLFV 207
Db 76 ----HATSDD-----LTNW-EDQFIAPKRN----- 97
QY 208 NSPAFTFAFLMEF-----LSNYHQADYLSTHYAEQGNHLEF 245
Db 98 DSGAFSGSMVVDYNNNTSGFFNDTIDPRQRCVAVIWTYNTPESEQYIS--YSLDGGYTFTE 155

```

```

246 AQRNLFAGVSPPEFKD-----SPRWQGTGISVLNTEIKQVYAD----- 284
Db 156 YQKPFVLAANSTQFEDPKVFWYEPQKIMT--AAKSQDYKIEIYSSDDLKSWKLESAPA 213
QY 285 -----GMQFEL-----SPIYHVAIDI-----FLKAY 306
Db 214 NEGFGLGYQECPLTEVTEQDPKSKSYWVFISINPGAPAGGSPNOYFVGSFNGTHFEAF 273
QY 307 GSAKRVNLEKEPPQSY--VQTVENMIMALISLSLDYNTPMFGDSWITDKNPRMAQFA-- 362
Db 274 DNQSRV---VDFGKDYALQTFN-----TDTYGSAL-GIAWAS--NWEYSAFVPT 319
QY 363 -----SWARVFPANQAIKYFA---TDGKQKAPNFKLSKALSNAAGFTFRSGWDKNATV 412
Db 320 NPWRSSMSLVKRFSLN--TEYQANPETELINLKABPILN--ISNAG-----PWSRFAFN 369
QY 413 MYLKASPPGEEHQAQPDNGTFFELFKGRNFTPDAGVVFVSGDEAIMK-----LRNWRQTR 467
Db 370 TTLTKANSYVNDLSNSTGTLEFEL-----VYAVNTTQTISKSVFADLSLWFKGLE 419
QY 468 -----IHSTLTLDNQNMVITKARQNKWETG-----NNLDVLTVTNPSYPN----- 507
Db 420 DPEEYLRMGFEVSGASSFFLDKRGNSKVKFKENPYFTNRMVNNQPFKSENDLSYKVKYGL 479
QY 508 LDHQRSVLFIN-----KKYFLVIDRAIGETAGNLGV 538
Db 480 LDQNIILELYFNDGDVVSVNTYFTMTGNALGSYNNMTGV 517
RESULT 2
SACC BACSU STANDARD; PRT; 677 AA.
AC P05656;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Levanase precursor (EC 3.2.1.65) (2,6-beta-D-fructan
DE fructanohydrolase).
GN SACC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87286401; PubMed=3112519;
RA Martin I., Debarbouille M., Ferrari E., Klier A., Rapoport G.;
RT "Characterization of the levanase gene of Bacillus subtilis which
RT shows homology to yeast invertase.";
RL Mol. Gen. Genet. 208:177-184(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=88067786; PubMed=3120151;
RA Schoergendorfer K., Schwab H., Lafferty R.M.;
RT "Nucleotide sequence of a cloned 2.5 kb PstI-EcoRI Bacillus subtilis
RT DNA fragment coding for levanase.";
RL Nucleic Acids Res. 15:9606-9606(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97286548; PubMed=9141695;
RA Parro V., San Roman M., Galindo I., Purnelle B., Bolotin A.,
RA Sorokin A., Mellado R.P.;
RT "A 23911 bp region of the Bacillus subtilis genome comprising genes
RT located upstream and downstream of the lev operon.";
RL Microbiology 143:1321-1326(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,

```

RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaterr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krohn S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
RA Pressean E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Iognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viarl A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256 (1997).
RN [5]
RP SEQUENCE OF 1-10 FROM N.A.
RX MEDLINE=90355183; PubMed=2117666;
RA Martin-Verstraete I., Debarbouille M., Klier A., Rapoport G.;
RT "Levanase operon of *Bacillus subtilis* includes a fructose-specific
RT phosphotransferase system regulating the expression of the operon.";
RL J. Mol. Biol. 214:657-671 (1990).
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 2,6-beta-D-
CC fructofuranosidic linkages in 2,6-beta-D-fructans (levans)
CC containing more than 3 fructose units
CC -1- INDUCTION: IN CONTRAST WITH LEVANSUCRASE AND SUCRASE, LEVANASE
CC SYNTHESIS IS NOT INDUCIBLE BY SUCROSE: NO INDUCER IS KNOWN FOR
CC LEVANASE.
CC -1- MISCELLANEOUS: LEVANASE CANNOT BE DETECTED IN THE WILD-TYPE
CC *B. SUBTILIS* BUT IS MOSTLY SECRETED INTO THE CULTURE MEDIUM BY SACL
CC MUTANTS, ESPECIALLY AT THE END OF THE EXPONENTIAL GROWTH PHASE.
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05649; CAA29137.1; ALT_INIT.
DR EMBL; Y00485; CAA68542.1; -
DR EMBL; X92868; CAA63465.1; -
DR EMBL; Z99117; CAB14645.1; -
DR EMBL; X56096; CAA39581.1; -
DR F01; A27286; A27286.
DR Subtilisin; BG10320; sacC.
DR InterPro; IPR001362; Glyco hydro 32.
DR Pfam; PF00251; Glyco.hydro_32; 1.
DR SMART; SM00640; Glyco_32; 1.
DR PROSITE; PS00609; GLYCOSYL HYDROL F32; 1.
KW Hydrolase; Glycosidase; Signal; Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 677 LEVANASE.
FT ACT SITE 49 BY SIMILARITY.
FT CONFLICT 658 658 Q -> L (IN REF. 2).
SQ SEQUENCE 677 AA; 75951 MW; 80FD6B0A5EE7F525 CRC64;

Query Match 3.4%; Score 120; DB 1; Length 677;
Best Local Similarity 18.8%; Pred. No. 1;
Matches 123; Conservative 83; Mismatches 216; Indels 232; Gaps 32;
QY 41 YSGLEKYNKVAAGNYDDAAKALLAYY---REKSKAREPDFSNAEKPADIQPIDKVTRE 97
DB 105 FSGSAVDKNTSGFQTKGKPLVAITYQREGHQVQSIAYSNDK-----GRWTWK 155
QY 98 MADKALVHQFQPHKGYGFDYGDKNQWMPVKDNEVRQLHRVKKWQAMA-----LVYH 152
DB 156 YAGNPVI-----PNPG-----KKDF-----RDPKVFWEKEKKWVLAAGDRILIY- 197
QY 153 ATGDEKVARWVVIQYSDWARKNPGLGLSDNDKFWVRLEVSDRVQSIPPTFSLFVNSEAF 212
DB 198 ---TSKNLKMOTY-----ASEFGDQGGSHGVW-----ECDFLFLFVDGNPN 237
QY 213 TPAPLMEF-----LNSYHQADYLSTHYVAEQG---NHLRF-EAQRNLEAGVSPEF 259
DB 238 QKKWQVQSVGNAVSGSGMGQYFVGDFDTHFKNENPPKVLMTDYGDFYAAVWSDI 297
QY 260 --KDSPR-----WRQTGISVLNTEIKKKQVYADGMQFELSPIYHVA 298
DB 298 PSTDSRLRLGLWMSNWOYANDVPTSPWSA--TSIPRELKLFKAFTEGVRVVQTPVKELET 355
QY 299 I-----DIFLKAYSAAKVNLE-KEPPOS-----YVQTVENMIM--- 331
DB 356 IRTGSKKKNLTTSPASHNVLAGOSDAYEINAEFKVSPGSAABFGFKVKTGENQFTKVG 415
QY 332 ----ALISISLPDYNTPMEGDSWITDKN---FRMAQFASWARVPFPAQAIKYFATDQK 383
DB 416 YDRNNAKLFVDRSESGNDTFNPAFTGKETAPLKPNGKVKLRFVDRSSVFEVFGNDGK 475
QY 384 GKA---PNFLSKAL---SNAG-----FYTPRSGWDKNATVMVLK 416
DB 476 VITDIILPDRSSKGLLYAANGVKVKSILTHPLKKVWGTTPPMSNMTGW---TTV--- 528
QY 417 ASPPGFHAQPDNGTELFELKGNFTPDAGVFVS-----GDEA 455
DB 529 -----NGTWADTIEGKQSGSDGDSFTLSSASGSDFTYESDITIKDNGRGAGA 576
QY 456 IMKLRNWRQTRIHSTLTLDNQNMWITKARQNKWETGNLNLVLTYNPSYPLNDHQRSVL 515
DB 577 LM-----FRSKDKAKGYLANVDKHDLVKFFKENGAAASVIAEYKTP-----I 620
QY 516 FINKKVL-----VIDRAITGEA-----TGNLGVH-WOLKEDSNPVF 550
DB 621 DVNKKYHLKTEABGRFKIYLDRLVIDAHSVFSEGFGLNWN---DATAVF 670
RESULT 3
C7AB_BACUA STANDARD; PRT; 1138 AA.
ID C7AB_BACUA
AC Q45707;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticidal crystal protein cry7AB (insecticidal delta-endotoxin
DE CryVIIA(b)) (Crystalline entomocidal protoxin) (130 kDa crystal
DE protein).
GN CRY7AB OR CRYVIIA(B).
OS *Bacillus thuringiensis* (subsp. *dakota*).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.
OX NCBI_TaxID=132268;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD511;
RA Payne J.M., Fu J.M.;
RT "Coleopteran-active *Bacillus thuringiensis* isolates and genes encoding
RT coleopteran-active toxins.";
RL Patent number US286486, 15-FEB-1994.
CC -1- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT
CC EPITHELIAL CELLS OF COLEOPTERA.
CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING

RT "Human apolipoprotein B: structure of carboxyl-terminal domains,
RT sites of gene expression, and chromosomal localization.";
RL Science 230:37-43(1985).
RN [10]
RP SEQUENCE OF 1-291 FROM N.A.
RX MEDLINE=86149325; PubMed=3513177;
RA Proter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V.,
RA Chen G.C., Kirsner S.W., McEnroe G., Kane J.P.;
RT "Isolation of a cDNA clone encoding the amino-terminal region of
RT human apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
RN [11]
RP SEQUENCE OF 1-1670 FROM N.A.
RX MEDLINE=86287319; PubMed=3461454;
RA Proter A.A., Hardman D.A., Sato K.Y., Schilling J.W.,
RA Yamanaka M., Hort Y.J., Hjerrild K.A., Chen G.C., Kane J.P.;
RT "Analysis of cDNA clones encoding the entire B-26 region of human
RT apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
RN [12]
RP PARTIAL SEQUENCE, AND IDENTIFICATION OF APO-B48.
RX MEDLINE=88018019; PubMed=3659919;
RA Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H.,
RA Silberman S.R., Cai S.-J., Deshpere J.P., Rosseneu M.,
RA Gotto A.M. Jr., Li W.-H., Chan L.;
RT "Apolipoprotein B-48 is the product of a messenger RNA with an organ-
RT specific in-frame stop codon.";
RL Science 238:363-366(1987).
RN [13]
RP DOMAINS.
RX MEDLINE=87039351; PubMed=3773997;
RA Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr.,
RA Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R.,
RA Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,
RA Levy-Wilson B., Scott J.;
RT "Complete protein sequence and identification of structural domains
RT of human apolipoprotein B.";
RL Nature 323:734-738(1986).
RN [14]
RP DOMAINS.
RA Yang C.-Y., Chen S.-H., Gianturco S.H., Bradley W.A., Sparrow J.T.,
RA Tanimura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,
RA Lee F.-S., Gu Z.-W., Gotto A.M. Jr., Chan L.;
RT "Sequence, structure, receptor-binding domains and internal repeats
RT of human apolipoprotein B-100.";
RL Nature 323:738-742(1986).
RN [15]
RP CALCIUM-BINDING DATA.
RX MEDLINE=86242245; PubMed=3087360;
RA Dashti N., Lee D.M., Mok T.;
RT "Apolipoprotein B is a calcium binding protein.";
RL Biochem. Biophys. Res. Commun. 137:493-499(1986).
RN [16]
RP PALMITOYLATION OF CYS-1112.
RX MEDLINE=20143590; PubMed=10679026;
RA Zhao Y., McCabe J.B., Vance J., Berthiaume L.G.;
RT "Palmitoylation of apolipoprotein B is required for proper
RT intracellular sorting and transport of cholesterol esters and
RT triglycerides.";
RL Mol. Biol. Cell 11:721-734(2000).
RN [17]
RP VARIANT SER-4338.
RX MEDLINE=91071750; PubMed=1979313;
RA Navajas M., Laurent A.-M., Moreel J.-F., Ragab A., Cambou J.-P.,
RA Cuny G., Cambien F., Roizes G.;
RT "Detection by denaturing gradient gel electrophoresis of a new
RT polymorphism in the apolipoprotein B gene.";
RL Hum. Genet. 86:91-93(1990).
RN [18]
RP VARIANT FDB GLN-3527.
RX MEDLINE=89098975; PubMed=2563166;
RA Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
RA McCarthy B.J.;

RT "Association between a specific apolipoprotein B mutation and
RT familial defective apolipoprotein B-100.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
RN [19]
RP VARIANT LEU-2739.
RX MEDLINE=91016974; PubMed=2216805;
RA Huang L.-S., Gavish D., Breslow J.L.;
RT "Sequence polymorphism in the human apoB gene at position 8344.";
RL Nucleic Acids Res. 18:5922-5922(1990).
RN [20]
RP VARIANT FDB CYS-3558.
RX MEDLINE=95190020; PubMed=7883971;
RA Pullinger C.R., Hennessy L.K., Chatterton J.E., Liu W., Love J.A.,
RA Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;
RT "Familial ligand-defective apolipoprotein B. Identification of a new
RT mutation that decreases LDL receptor binding affinity.";
RL J. Clin. Invest. 95:1225-1234(1995).
RN [21]
RP VARIANTS LEU-1437; SER-1914; LYS-2566; THR-3121; ALA-3945; MET-4128
RP AND THR-4481.
RX MEDLINE=97044521; PubMed=8889592;
RA Poirier O., Ricard S., Behague I., Souriau C., Evans A.E.,
RA Arveiler D., Marques-Vidal P., Luc G., Roizes G., Cambien F.;
RT "Detection of new variants in the apolipoprotein B (Apo B) gene by
RT PCR-SSCP.";
RL Hum. Mutat. 8:282-285(1996).
RN [22]
RP VARIANTS FDB GLN-3527 AND CYS-3558.
RX MEDLINE=97403938; PubMed=9259199;
RA Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,
RA Krempf M., Giraudet P., Junien C., Boileau C.;
RT "Familial ligand-defective apolipoprotein B-100: simultaneous
RT detection of the ARG3500->GLN and ARG3531->CYS mutations in a
RT French population.";
RL Hum. Mutat. 10:160-163(1997).
RN [23]
RP VARIANTS SER-1914; ARG-1923; LEU-2739; ASP-3319; THR-3427; GLN-3432
RP AND ILE-3921.
RX MEDLINE=98141125; PubMed=9490296;
RA Lerer T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;
RT "Screening for mutations of the apolipoprotein B gene causing
RT hypercholesterolemia.";
RL Hum. Genet. 102:44-49(1998).
CC -!- FUNCTION: APOLOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF
CC CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION SIGNAL
CC FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL PARTICLES BY
CC THE APOB/E RECEPTOR.
Query Match 3.3%; Score 116; DB 1; Length 4563;
Best Local Similarity 16.3%; Pred. No. 29;
Matches 104; Conservative 94; Mismatches 197; Indels 244; Gaps 26;
QY 218 MEFLNSY-----HQADY-----LSTHYAEQGNHRLFEA-----QRNL----- 250
DB 3068 IDFLNLYALFLSPSAQASQVARSARPNQYKQNFSGNENIMEAHVINGEANLDPLN 3127
QY 251 -----FAGVSFFPEFKDSPRWQTG----- 269
DB 3128 IPLTIPEMELPYTIITTPPLKDFSLWEKTKQFELTKTQSFDSLVKQYKKNKRRHSIT 3187
QY 270 ---ISVUNTEIKQVADGMQFELSPYHVAADIFLKAYSAGK-----RVNLEK---EFFQ 320
DB 3188 NPLAVLCEPISQSIKSFDRHFEKN---RNNALDFVTKSYNETKIKIPDKYKAKSHDELPR 3244
QY 321 SVYQVTVENIMMALISISLPDYNTPMGDSWITDKFRMAQF-----ASMARVFPANQAIKY 376
DB 3245 TP-----QIEGYTVPVV-----NVEVSPFTIEMSAFGYVFPKAVSMPS 3282
QY 377 FATDGKQGGKAPNF-----LSKALSNAGF-YT 401
DB 3283 FSLGSDVRVPSYTLTLPSELPLVHLVPRNLKLSLPHFKELCTISHIFTAMGNITYDFS 3342
QY 402 FRSGWKNATVVLKASPFGEFHAQPDNGTFELFIKGRNFTPDAGVVFVSGDEAI----- 456

Db 3343 FKSS -----VITLNTAELFNQSD--IVAHLSSSSSVIDALQYKLEGTTLTRKRG 3392
QY 457 -----MKLRNVQRTRIHSTLTLDNONAVITKARONKWE-----TGNLDVL 498
Db 3393 LKLATALSLSNKFVEGSHNSTVSLTTKNNEVSVAKTTKAEIPILRMNFQELNGNTKSKP 3452
QY 499 TVTNPSYPLNDHQRSVLFNKK-----YELVIDRAIGEATGNL----- 536
Db 3453 TVSSSMFYKDFNSMLYSTAGAVDHKLSLESLSYFSIESSTKGDVGVLSREYSOT 3512
QY 537 -----GVH-----WOLKEDSNPVFDKTKNRVYTYTYRDG--NNLM 568
Db 3513 IASEANTYLSKSTRSSVKLGTSKIDDIWNLEVENFAGEATLQRIYSIWEHSTKHQLQ 3572
QY 569 IOSL---NADRTS---LNBEKQSVVYNKELKRPAPVFEKP-----KKNAGTQN----- 612
Db 3573 BEGLFTNGEHTSKATLELSPWQMSALVQVHASQSPSPHDFDQGVFALNANTKNQIR 3632
QY 613 FVSIYVPYDQKAPESISRENGNDPEKGLNLTITNG 651
Db 3633 WKNEVRIHSGSPQSVEL-----SNDQEKALHDIAGSLEG 3667

RESULT 6
Y140 MYCPN
ID Y140 MYCPN STANDARD; PRT; 1113 AA.
AC P75033;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical ATP-binding protein MG140 homolog (E07_orf1113).
GN MEN153 OR MP001.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
EX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SIMILARITY: BELONGS TO THE DN42/NAW7 HELICASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000001; AAB95649.1; --
DR PIR; S73327; S73327.
KW Hypothetical protein; ATP-binding; Helicase; Complete proteome.
FT NP_BIND 313 320 ATP (POTENTIAL).
SQ SEQUENCE 1113 AA; 130333 MW; 48A3337EB0E81A40 CRC64;

Query Match 3.3%; Score 115.5; DB 1; Length 1113;
Best Local Similarity 16.9%; Pred. No. 4.2; Mismatches 241; Gaps 29;
Matches 118; Conservative 102;

QY 15 IALSSGNILAOSSSITRKD-----FDHINLEYSGLEKVNKVAAGNYDDAAKALLAY 66
Db 330 IALNKKVLFVSEKVTALDVVYNRLGSPFHIAL-----FNASVAS-----EKRFPYQ 377
QY 67 TREKSKAREPDSNA-----EKPADTRQPIDKVTREMAKALVHQFQPHKGYFYDGKDI 122
Db 378 FADYESFTDNFSKDLVNEMEPVFDQWVDKILSEFTNLQNIYDTQINSGNQSYF-KEI 436

QY 123 NQMMPVKDNEVRWQLHRVKKWQAMALVYHATGDEKYAREWYQYSD--WARKNPLGLSQ 180
Db 437 -LSSFPILD-----VSYIKIKEHDFDEWVRVFSQVWLEKHLTYLA- 477
QY 181 DNDKEFVWRPLEYSDRVQSLPPTFSLFVNSPAFTAPLME-----FLNSVHQOQD-- 229
Db 478 -----FKAELS KWQNIDNEVAL-----KDLLEKKNIRVLCYVLDYFEQNNSI 521
QY 230 -----YLSHYABQGNHRLFEAQRLNLFAGVSPFEDKSPRQGTGISVLNTEIKQV 281
Db 522 IKPKRVLLYTPT---ERGQQLHQDQVAKYNSLQRFKSAKFETIKLNANK----- 572
QY 282 YADGMQFELSPLHYHVAIDIFLKAYGSAKRVNL---EKEF---POSYQTVENMIMA--- 332
Db 573 ---LAQNAKPPF---FSWFIQTHAQTLLENLVQTKOLVRKAKOSYLSKIQYVVVCKR 624
QY 333 -----LISISLPDYNTPMFGDSWITDKNFRMAQ-----FAS 363
Db 625 ILKATILANFFELYQTNKNELLDICREAKNPVLKEITWFKKNEFALLSKLPVHIMTFES 684
QY 364 WARVFPANQAI-----KY-FATDGKQKAPNF----- 389
Db 685 AALLTENQRRLYDVYVIDEASQVYLERAIPILYRGAKYIAGDTKOLKPSNFFQARAED 744
QY 390 -----LSKALSNAGFYTFERSGDKNATVMVLKASPPGEFHAQPDNGTELFIKGN 440
Db 745 VDEEFEDGNVEAAVHSTSLHFLKNRSRLITLL-----KFHYRSDSANLIAFTNNR- 795
QY 441 FTPDAGVFVYSGDEAIMKLRNMYRTRIHSITLTLDNQNMVITKARQNKWETGNL---D 496
Db 796 -----LYNNELIFM-----NKATAQKQFVIVHDVIDGIWNNNNNQEAR 835
QY 497 VLTYNPSYPLNDHQRS---VLEINKYFLVIDRAIGBATGNLGVHWQLKEDSNPVDPKTK 554
Db 836 VVQRLEQLTQTAEYQKSLGVICFNKNQAEILIE-----YMDKQNNPLINENR 882
QY 555 NRV-----YTYRDGNLMLQSLNADRT 577
Db 883 DRVNAQGEVGLFVKNIENVOGDERDIIFSLGYDRS 919

RESULT 7
SS22 YEAST
ID SS22 YEAST STANDARD; PRT; 1314 AA.
AC P25390;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Serine/threonine protein kinase SSK22 (EC 2.7.-.-).
GN SSK22 OR YCR073C OR YCR73C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Ballesta J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,
RA Sanz E.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95350642; PubMed=7624781;
RA Maeda T., Takekawa M., Saito H.;
RT "Activation of yeast Pbs2 MAPKK by MAPKKs or by binding of an SH3-
RT containing osmosensor."
RL Science 269:554-558(1995).
CC -!- FUNCTION: KINASE INVOLVED IN A SIGNAL TRANSDUCTION PATHWAY THAT IS
CC ACTIVATED BY CHANGES IN THE OSMOLARITY OF THE EXTRACELLULAR
CC ENVIRONMENT. ACTIVATES THE PBS2 MAP KINASE KINASE BY
CC PHOSPHORYLATION.
CC -!- SUBUNIT: INTERACTS WITH BY SGK1.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAP KINASE KINASE SUBFAMILY.


```
FT VARIANT 1050 1050 L -> F (IN STRAIN PS178D4).
FT CONFLICT 1094 1118 GHRITQACTWKRONSVRSWKR -> EDTEYKLRVRGK
FT GKGTVYVQHGEE (IN REF. 21).
SQ SEQUENCE 1167 AA; 131658 MW; 47ALED4EBDC645 CRC64;

Query Match 3.3%; Score 114.5; DB 1; Length 1167;
Best Local Similarity 19.7%; Pred. No. 5.3;
Matches 103; Conservative 62; Mismatches 170; Indels 189; Gaps 24;

QY 249 NLFAGVSPPEF--KDSPRWRQTGISVLNTEIKQV-----YADGNQFELS 291
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 94 NMFVDIFPKLFGSGSQNAQFFELIEKVELVDEFRNFTLNLLNYLDNGQTALS 153
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 292 PIYHVAADIFLKAGSRAKVNLEKFPQSVQTVENMI-----MALISISLPDYN 342
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 154 HFQNDVQIAI---COGEPGLMLD-QTPACTPTDHLISVRESFKDARTIETALPHK 209
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 343 TPFMGDSWITDKNFRMAQFASWARVFPANQAIKYPATGKQKQAFNLSKALS-AG 398
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 210 NPMLS-----TNDNTPDFNSDTVLLTLPWYTTGATNLNLHQQ 247
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 399 FYTPRSGWCK-----NATVMVLKASPPGEFHAQPDNGTFELIKGRNFTPDAGVFY 450
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 248 YIQFAERKSVNYDESFINQTKVOLRRI--QDYSTTVSTTPEKFKPTLN-----P 296
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 451 SGDEAIMKLRWYQTRIHST-----LTLDN-----QNMVITKARQKWE-- 490
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297 SNKESVKNYKRVSRMTLQSLDIAATWPTLDNVNVPNSVDIQLDQTRLVFSDVAGPWE 356
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 491 ---TCNNLDVLTYTNP-----SYNLDHQ-----RSV-----LF 516
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 357 DNITSNIIDVLTPIGTIGFQESSDLRKFTYPIELQSMQPHGQVNSKSVCHYSDGLK 416
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 517 INKKYFLVIDRAIGATGNLGVHWQKEDSN-----PVFDKTRVYTYTD 563
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 417 LNYK-----NKTITAGVSN-----DESNQNNKHNGPVINSPITDINVSQSLD 464
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 564 GNNLMIQ-----SLNADRTSLNEEKGVSVYNNKELKRPDAFVPEKKNAGTQ- 611
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 465 LNSVMVNGGQKVTGCSPLSSNGSNNAALPNQKINVISVQSN-----DKPEKHADTYR 518
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 612 ---NFVSIYVPYD-----GQKAPETISIRENKGNDFPKG 641
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 519 KWGYMSSHIPYDLVPENVIGIDPDTKQPSLLL---KGPPAEKG 559
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
EX5B_HAEIN STANDARD; PRT; 1211 AA.
AC P45157;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Exodeoxyribonuclease v beta chain (EC 3.1.11.5).
GN RECB OR H1321.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Keiley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae"
```

```
RT Rd.";
RL Science 269:496-512 (1995).
CC !- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE
CC UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
CC STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP
CC (BY SIMILARITY).
CC !- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
CC ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
CC phosphooligonucleotides.
CC !- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD
CC (BY SIMILARITY).
CC !- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVRD SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32811; AAC22966.1; -.
DR PIR; D64116; D64116.
DR HSSP; P09980; 1UAA.
DR TIGR; H11321; -.
DR InterPro; IPR004586; RecB.
DR InterPro; IPR000212; UvrD-helicase.
DR Pfam; PF00580; UvrD-helicase; 1.
DR TIGRfam; TIGR00609; recB; 1.
KW Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
KW DNA repair; Complete proteome.
FT NP_BIND 22 29 ATP (POTENTIAL)
SQ SEQUENCE 1211 AA; 139857 MW; 37B76430651PAD68 CRC64;

Query Match 3.3%; Score 114.5; DB 1; Length 1211;
Best Local Similarity 20.1%; Pred. No. 5.6;
Matches 102; Conservative 67; Mismatches 190; Indels 149; Gaps 23;

QY 53 AGNYDDAAKALLAYVREKSKAREPDSNAEKPADIR-----QPIDKVTREMDKALVHQ- 107
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 725 AATLHSEAAALLSWFEKQIQ-----GEARQEQAIRLESEKOLVIVITIHKSGLEYDILV 778
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 108 -----OPHKGGYGYFYKQINQWMPVKD-----NEVRW---QLHRVK 142
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 779 WLPFLAAPSKPSKXYINIIYSKERDETLNDIENRNALCEETFAEELRLIYVALTRAK 838
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 WQQAALVYHATGDEKY-AREWVYQYSDWARKNPLGLSQDNDKFWVRPLE-VSDRVQ--- 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 839 YQMAFAL--PAQFDKKNALHYVLVLSQGEIGKEINLSDSKDTETLLQTFEKKMQDNVEICT 896
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 198 ----SLPTFSLVNSPAPTAFALMEFLNSVHQADYLSHYAEQGNHRLFEAQNLF-- 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 897 KPNLEALPTLSINTKNDDFKAS---EFTGNIEQDWRITSTFIEQAHR- ---QNYFTE 948
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 252 -AGVSFPFEFKOS-----PRMRTGIS-----V 272
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 949 SAGKKHAFDDAKDYDSQNAIEISTALLNENESNILDLPGRKQVGTALHRRHFCYFSDL 1008
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 273 LNT-BIKK-----QVYADGMQFELSPIYHVAVIDFLKAYGSAKRVNLEKEFPQSV 323
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1009 ANTEBIDKLRSQQLDDETFETESLQNLQWQISHTPLSLNEIGIALADLANKDCKEMP-FYL 1067
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 324 QTVENMIMALISISLPDNTYNTMFGDSWITDKNFRMAQFASWAR-----VFPANQAIK 378
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1068 AIREH-----FDVEARNHTLKAHHLPSPESLQFEQIQGMVRSGLDILVFRHNG--KY 1118
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 379 TDGKQKAPNLSKALSNA-----FYTPRSGWKNATVMVLKASPP 420
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1119 VDYKS---NFLGSLTADVNOEALKKEMLSHYDWOYLIYTL-----ALHRYLSQSV 1167
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 421 GEFHAQPDNGTFELIKGRNFTPDAGVF 448
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

SIGNAL	1	?	POTENTIAL
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			
91			
92			
93			
94			
95			
96			
97			
98			
99			
100			

QY 548 PVFDKTKNRVY-----TYRDGNMLIQSLNADRTSLNEBEGKVSYY----- 590
Db 1114 -----EARNYTLIVKHALTLEPVNAIV--LIIGNYTYLTDBNGVTFYTAPTKLSDEI 1165
QY 591 -----NKLKRPAPVFFKPKKNAQTQNFVSIVPYDG-- 622
Db 1166 TVIVKKNFNTLEKTFQITVSRPEITEEDNEPKLMSPEANA---TIVSMESEGGV 1222
QY 623 QKAPISIREKNGDNFEKGLNLTITNGKQQLVLP 659
Db 1223 KKTIVTEITIN-----GTANETATI-----WVFP 1247

RESULT 12

POLG_EC23W STANDARD; PRT; 2179 AA.
AC 073556;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Coat protein VP0 (PIAB); Coat protein VP3 (PIA); Coat protein VP1 (PID); Core protein 2A; Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VP0 (P3B); Poliovirus 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
DE Echovirus 23 (strain Williamson) (Human parechovirus 2).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Parechovirus.
OC NCBI_TaxID=122962;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99036025; PubMed=9820139;
RA Ghazi F., Hughes P.J., Hyypiae T., Stanway G.;
RT "Molecular analysis of human parechovirus 2 (echovirus 23).";
RL J. Gen. Virol. 79:2641-2650(1998).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES
CC AT CERTAIN Q/G SITES IN THE POLYPROTEIN. IT IS A CYSTEINE
CC PROTEASE.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-Gly bond in the
CC poliovirus polyprotein. In other picornavirus reactions Glu may be
CC substituted for Gln, and Ser or Thr for Gly.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP0, VP1,
CC AND VP3.
CC -!- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC ALL CLEAVAGES ARE CATALYZED BY P3C.
CC -!- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AJ005695; CAA06679.1; -.
CC MEROPS; C03.023; -.
DR InterPro; IPR007053; NC.
DR InterPro; IPR000605; RNA helicase.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_Psvir.
DR Pfam; PF04970; NC; 1.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA helicase; 1.
KW Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 289 COAT PROTEIN VP0.
FT CHAIN 290 542 COAT PROTEIN VP3.
FT CHAIN 543 775 COAT PROTEIN VP1.

FT CHAIN 776 922 CORE PROTEIN P2A.
FT CHAIN 923 1044 CORE PROTEIN P2B.
FT CHAIN 1045 1373 CORE PROTEIN P2C.
FT CHAIN 1374 1490 CORE PROTEIN P3A.
FT CHAIN 1491 1510 GENOME-LINKED PROTEIN VP0.
FT CHAIN 1511 1710 PICORNAIN 3C.
FT CHAIN 1711 2179 RNA-DIRECTED RNA POLYMERASE.
FT SITE 763 765 CELL ATTACHMENT SITE (POTENTIAL).
FT ACT_SITE 1669 1669 PROTEASE 3C (POTENTIAL).
FT ACT_SITE 1687 1687 PROTEASE 3C (POTENTIAL).
SQ SEQUENCE 2179 AA; 245872 MW; 96803COBB8856664 CRC64;

Query Match

Best Local Similarity 3.2%; Score 113; DB 1; Length 2179;

Matches 140; Conservative 89; Mismatches 261; Indels 210; Gaps 36;

QY 4 KIFKRIIVFAVIALSSGNIQAQSSITRKDFPHINLEYSGLKYNKAVAGNYDAAKAL 63
Db 953 KILVRILCYMVLVYCHKPNILTTACLSTLLVMDVTS--SSVLSPSCALMQCLMDGDEVK 1010
QY 64 LAYYREKSKAREPDFSNAEKPADIRQPDVKVTREMA-----KALVHQFQPHKGYGYFDY 118
Db 1011 AEVVAE-----SMSNTDD-----DEIKEQICDVTYTKQILSNQGPFGF----- 1050
QY 119 GKIDINQMWPVKNDEVRWQLHRVKW-----QAMALVYHATGDEKYAREVWYQYSDWA 171
Db 1051 -----NEISTAFRHIDMWIQTLLKIKDWLVSFKPVSVEKRAVEL----- 1090
QY 172 RKNPLGLSQDNKDFWRPLEVSDRVQSLPPTSLFVNSPAPFAELMEFLNSYHQOAYL 231
Db 1091 -----ERNKEHVCISILDYAS-----DIIVKSKQDTKMKTKQEFYQRYN---DCL 1130
QY 232 STHYAEQGNHRLFEAQNLFAGVSPFEPKDSR-----WRQ-----T 268
Db 1131 SKFEKIMA--MCFRSCHNSISNTVYRLPQELARIENRMTATQNDLIRVEPIGIWIQGEPCQ 1188
QY 269 GISVLNTRIKQOVAD--GMQFELSPIYHVAIDIFLKAYGSAKRVNLEKEFPQSVYQVTE 327
Db 1189 GKSFLTHTLSKQLQKTCGLQ-----GIYTNPTASEFMDGYDN-QDIHLIDDLGQTRKERDI 1243
QY 328 NMIMALISISLPDYNTPMFGDSWITDKFRMAQFASWARVFAQNAIKYFATDQKQKAP 387
Db 1244 EMLCNCIS--SDPDI-----VPMALHEEKGFYTSKLV-----ATTNK-----P 1281
QY 388 NFLSKALSNAAGFYTFERSGMDKNATVMVLA-----SPPGEFH---AQPDMGTFFELFKGRN 440
Db 1282 DFSSTVLLDSGALRRFPY-----IMHFAAHKYSKGLNVSQAMPHMSTGECWEVSKN 1336
QY 441 FTFDAGVFVYSGDEAIMKLRNWRQTRIHSITLTLDNQNNVITKARONKW-----ETGNN 494
Db 1337 -----GRDWETLKLK-----ELIDKITVDYKERI-----ANYNTWKQLEDDQLDD 1377
QY 495 L-DVLTYYNPSYPN-----LDHQRSVLFINKKYFL-----VIDRAIGEATGNLGVHWQ 541
Db 1378 LDDAVSYTKHNPYDAIPYIDEYLNIEMLSTLIEQMEAFIEPKPSVFKCFASRVGD-----K 1432
QY 542 LKEDSNPFV-----DKTKNRVYTYTRDGNMLTQSLNADRTSL-----NEEBEGKVS 587
Db 1433 IKAASREVVKWFSFKLSML--NFVERNKAMLTVVSVAITSAIGILLVITKIFKESKDE 1490
QY 588 YVYNKELKRPAPVFFKPKKNAGT-----QNFVSIVVPYDQG 623
Db 1491 RAYN-----PTLPVAKPK---GTFPVVSQREFKNEAPYDQG 1522

RESULT 13

ID INVI_YEAST STANDARD; PRT; 532 AA.
AC P10594;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Invertase 1 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 1)

```
(Saccharase).
DE SUC1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RN SEQUENCE OF 1-73 FROM N.A.
RP MEDLINE=88216256; PubMed=2835632;
RX Hohmann S.; Gozalbo D.;
RA "Structural analysis of the 5' regions of yeast SUC genes revealed
RT analogous palindromes in SUC, MAL and GAL.";
RL Mol. Gen. Genet. 211:446-454(1988).
[2]
RN SEQUENCE FROM N.A.
RP Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RL -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X07570; CAA30457.1; -
CC SGR; S27372; S27372.
CC PIR; L0002143; SUC1.
CC InterPro; IPR001362; Glyco_hydro_32.
CC Pfam; PF00251; Glyco_hydro_32; 1.
CC SMART; SM00640; Glyco_32; 1.
CC PROSITE; PS00609; GLYCOSYL HYDROL F32; 1.
CC Hydrolase; Glycosidase; Glycoprotein; Multigene family; Signal.
CC SIGNAL 1 19
CC CHAIN 20 532
CC ACT_SITE 42 42
CC CARBOHYD 23 23
CC CARBOHYD 64 64
CC CARBOHYD 111 111
CC CARBOHYD 112 112
CC CARBOHYD 118 118
CC CARBOHYD 165 165
CC CARBOHYD 275 275
CC CARBOHYD 356 356
CC CARBOHYD 369 369
CC CARBOHYD 384 384
CC CARBOHYD 398 398
CC CARBOHYD 512 512
CC SEQUENCE 532 AA; 60570 MW; 3280EF260EA01606 CRC64;
Query Match 3.2%; Score 112; DB 1; Length 532;
Best Local Similarity 18.6%; Pred. No. 2,6;
Matches 107; Conservative 83; Mismatches 176; Indels 210; Gaps 31;
QY 94 VTREMAKALVHQFQPHKGY-----GYFDYKIDINQMW-PVKDNEVRQLHRVQWQAM 147
DB 21 MTNETSDRPLVH-FTFNKGWMDPNGLWDYDAKEGKWHLYFQYNPDVTWGL--PLFWG-- 75
QY 148 ALVYHATGDEKAREWVQYSDWARKNPLGLSQDNDKFWWRPLEVSDRVQSPPFESLFV 207
DB 76 ----HATSDD-----LTHW-----QD-----EVAIAPRK-- 97
QY 208 NSPAPFLPMBF-----LNSVHQADYLSHYAEOGNHRLFE 245
DB 98 DSGAYSGSNVIDYNNSTGFFNDTIDPRQCVAIWYNTPESEQVIS--YSLDGGYTFTE 155
QY 246 AQRNLFAGVSPFEPKD-----SPRWRQTGISVLNTEIKQVYAD----- 284
DB 156 YQKNPVLAAANSQFRDPKVFWEPSKKWTMT--AAKSQDYKIEIYSSDDLKSKWLESAPA 213
```

```
285 -----GMQFEL-----SPIYHVAADI-----FLKAY 306
214 NEFGLGYQYECPLIEVPSEQDPKSHWYMFISINFAPAGGSFNQYFVGSFNGHHEAF 273
307 GSAKRVNLEKEFPQSY--VOTVENMIMALISISLDPYNTPMFGDSWITDKNPEMAOF--- 361
274 DNQSRV---VDFGKYVALQTFN-----TPTYGSAL-GIAWAS--NWEYSAPVPS 319
362 ----ASWARVFPANQAIKYFA---TDGKQKAPNLSKALSNAAGFYTFERSGWDKNATVMV 414
320 NPWRSSMSLVRFPSLNTYQANPETELINLKAEPILN--ISSAG-----PWSRFATNTT 371
415 LKASPPGEPHAQPDNGTFFLFIKGRNFTPDAGVFVYSGDEAIMK-----LRNRYOTR-- 467
372 LTKANSYNVDLSNSTGTLEFEL-----VYAVNTTQTISKSVFADLSLWFKGLEDP 421
468 -----IHSTLTDNONNVITKARONKWEFG---NLLDVLTYTNPSYPN---LD 509
422 EBYLRMGFEVSASSFELDRGNSKVFEKVPFTVRMSVNNQPFKSENLSYIKVYGLLD 481
510 HORSVLFIN-----KKYFLVIDRAIGBATGNLGV 538
482 QNILELYFNDGDWSTNTVEMTTGNALGSVNMVTGV 517
RESULT 14
INV4_YEAST STANDARD; PRT; 532 AA.
AC P10596;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE Invertase 4 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 4)
DE (Saccharase).
OS SUC4.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RN SEQUENCE OF 1-73 FROM N.A.
RP MEDLINE=88216256; PubMed=2835632;
RX Hohmann S.; Gozalbo D.;
RA "Structural analysis of the 5' regions of yeast SUC genes revealed
RT analogous palindromes in SUC, MAL and GAL.";
RL Mol. Gen. Genet. 211:446-454(1988).
[2]
RN SEQUENCE FROM N.A.
RP Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RL -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X07570; CAA30457.1; -
CC SGR; S27372; S27372.
CC PIR; L0002143; SUC4.
CC InterPro; IPR001362; Glyco_hydro_32.
CC Pfam; PF00251; Glyco_hydro_32; 1.
CC SMART; SM00640; Glyco_32; 1.
CC PROSITE; PS00609; GLYCOSYL HYDROL F32; 1.
CC Hydrolase; Glycosidase; Glycoprotein; Multigene family; Signal.
CC SIGNAL 1 19
CC CHAIN 20 532
CC ACT_SITE 42 42
CC CARBOHYD 23 23
CC CARBOHYD 64 64
CC CARBOHYD 111 111
CC CARBOHYD 112 112
CC CARBOHYD 118 118
CC CARBOHYD 165 165
CC CARBOHYD 275 275
CC CARBOHYD 356 356
CC CARBOHYD 369 369
CC CARBOHYD 384 384
CC CARBOHYD 398 398
CC CARBOHYD 512 512
CC SEQUENCE 532 AA; 60570 MW; 3280EF260EA01606 CRC64;
```



```

QY 411 TVMVLKASPPGEFHAQPDNGTFELFIKGRNFTPDAGVFVYSGDEAIMKLENWYRQTRIHS 470
Db 550 RVRVR-----YATNVSCQFNVI----- 567
QY 471 TLTLDNONNVITKARONKWT-CNNLDVLTNTNPSYENLDHQRSVLFINKKYFLVIDRAI 529
Db 568 -----NDKITLQRFQNTVETIGEGKD-LTYGSFGY--IEYSTTIQFPDKHPKITLH--L 617
QY 530 GEATGNLGVHWQLKE-----DSNPVFDKXNRVYTTYRDGNNLM---IQSLNADRTS 578
Db 618 SDSLNNSSFYVDSIEFIPVDVNVYDEKEKLEKAQKAVNTLFTEGRNALQKDVTIDYKVDQVS 677
QY 579 L 579
Db 678 I 678

```

Search completed: July 28, 2003, 20:50:14
Job time : 31 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 20:46:28 ; Search time 46 seconds
(without alignments)
1377.721 Million cell updates/sec

Title: US-09-802-285A-2

Perfect score: 3494

Sequence: 1 MTTKIFKRIIVFAVIALSSG.....KGKLNLTITNGKQLVLVP 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3443	98.5	659	JC4910	heparin-sulfate ly
2	189	5.4	633	A97309	hypothetical prote
3	176	5.0	546	F95038	conserved hypothet
4	160	4.6	1183	F90559	poly(beta-D-mannur
5	154	4.4	713	JC5870	hypothetical prote
6	139	4.0	561	T36084	p60-related protei
7	139	3.9	604	AB1695	hypothetical prote
8	137.5	3.9	604	T44835	hypothetical prote
9	127	3.6	560	E96554	beta-fructofuranos
10	125.5	3.6	672	AC0422	hypothetical prote
11	124.5	3.6	1883	G82875	hypothetical prote
12	123	3.5	532	IFBY	probable sulfatase
13	120.5	3.4	625	H90533	hypothetical prote
14	120.5	3.4	970	E70533	lipoprotein [impor
15	120	3.4	1272	C90593	levanase (EC 3.2.1
16	119	3.4	776	E90591	RF1 protein - yeas
17	118.5	3.4	677	A27286	toxin-like outer m
18	117.5	3.4	995	S07353	fibronectin-bindin
19	117.5	3.4	2902	C71953	apolipoprotein B-1
20	116	3.3	1117	S33851	MG140 homolog - My
21	116	3.3	4563	LPRUB	probable membrane
22	115.5	3.3	1113	F73327	hemoglobin-binding
23	115.5	3.3	1314	S19488	alpha-amylase (EC
24	115	3.3	953	B64083	alkaline amylopul
25	115	3.3	1256	G97902	probable flagellin
26	115	3.3	1280	E95031	exodeoxyribonuclea
27	114.5	3.3	750	G81361	type IIS restricti
28	114.5	3.3	1211	D64116	
29	114.5	3.3	1279	E64709	

ALIGNMENTS

RESULT 1

JC4910

heparin-sulfate lyase (EC 4.2.2.8) - Flavobacterium heparinum

N;Alternate names: heparin-sulfate eliminase; heparinase III

C;Species: Flavobacterium heparinum

C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 03-Jun-2002

C;Accession: JC4910

R;Godavarti, R.; Davis, M.; Venkataraman, G.; Cooney, C.; Langer, R.; Sasisekharan, R.

Biochem. Biophys. Res. Commun. 225, 751-758, 1996

A;Title: Heparinase III from Flavobacterium heparinum: Cloning and recombinant expression.

A;Reference number: JC4910; MUID:96374394; PMID:8780685

A;Accession: JC4910

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-659 <GD>

A;Note: the authors translated the codon ACC for residue 313 as Asn

C;Comment: This enzyme cleaves heparan sulfate-rich regions of acidic polysaccharides.

C;Superfamily: Flavobacterium heparinum heparitin-sulfate lyase

C;Keywords: carbon-oxygen lyase

Query Match 98.5%; Score 3443; DB 2; Length 659;

Best Local Similarity 98.8%; Pred. No. 8.5e-237;

Matches 651; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 MTTKIFKRIIVFAVIALSSGNILAQSSSITRKDFDHINLEYSGLKVNKAVAGNYDDAA 60

Db 1 MTTKIFKRIIVFAVIALSSGNILAQSSSITRKDFDHINLEYSGLKVNKAVAGNYDDAA 60

QY 61 KALLAYYREKSKAREPDFSNAEKPADIRQPIDKVTREMAADKALVHOFQPHKGYGFDYDK 120

Db 61 KALLAYYREKSKAREPDFSNAEKPADIRQPIDKVTREMAADKALVHOFQPHKGYGFDYDK 120

QY 121 DINQWMPVKDNEVQQLHRVQWQAMALVYHATGDEKYAREWVVOYSDWARKNPLGLSQ 180

Db 121 DINQWMPVKDNEVQQLHRVQWQAMALVYHATGDEKYAREWVVOYSDWARKNPLGLSQ 180

QY 181 DNDKFWVRPLEVSDRVQSLPPTFSLFVNSPAFTPAFLMEFLNSYHQADYLSHYAEQGN 240

Db 181 DNDKFWVRPLEVSDRVQSLPPTFSLFVNSPAFTPAFLMEFLNSYHQADYLSHYAEQGN 240

QY 241 HRLFAQRNLFAGVSFPFKSPRWRQTGISVLNTEIKQVYADGQMFELSPYHVAID 300

Db 241 HRLFAQRNLFAGVSFPFKSPRWRQTGISVLNTEIKQVYADGQMFELSPYHVAID 300

QY 301 IFLKAYGSAKRVNLEKEPQSVQTVENIMIALISISLPDYNTPMGDSWITDKNFRMAQ 360

Db 301 IFLKAYGSAKRVNLEKEPQSVQTVENIMIALISISLPDYNTPMGDSWITDKNFRMAQ 360

QY 361 FASWARVFPANQAIKYFATDGKQKQKAPNFKLSKALSNAGFYTFRSGWKNATVWLKASPP 420

Db 361 FASWARVFPANQAIKYFATDGKQKQKAPNFKLSKALSNAGFYTFRSGWKNATVWLKASPP 420

T16G12.1 protein -
protein T16G12.1 [
beta-fructofuranos
beta-fructofuranos
pitrilysin (EC 3.4
toxR-activated lip
toxR-activated gen
pyrolysin (EC 3.4.
hypothetical prote
hypothetical prote
call wall-associat
alkaline phosphata
hypothetical prote
lipoprotein [impor
hexon protein - hu
hypothetical prote

30 114 3.3 885 2 S42841
31 114 3.3 899 2 H88572
32 112 3.2 532 2 S27373
33 112 3.2 532 2 S27372
34 111.5 3.2 962 2 AC0125
35 111.5 3.2 1002 2 T09438
36 111.5 3.2 1013 2 B82276
37 111.5 3.2 1398 2 T28159
38 109.5 3.1 808 2 A82877
39 109.5 3.1 1465 2 A70199
40 109.5 3.1 2334 2 S32920
41 109 3.1 634 2 F97199
42 109 3.1 712 2 F97012
43 109 3.1 789 2 G90587
44 109 3.1 952 1 HXAD5
45 108.5 3.1 554 2 A70120

QY 421 GEFHAPDNGTGFELFIKGRNFTPDAGVYVYSGDEAIMKLRNWRQTRIHSITLTDNQNV 480
Db |||||
QY 421 GEFHAPDNGTGFELFIKGRNFTPDAGVYVYSGDEAIMKLRNWRQTRIHSITLTDNQNV 480
Db |||||
QY 481 ITKARQNKWETGNNLDVLTYNPSYPNLDHORSVLFINKKYFLVIDRAIGATGNLGVHW 540
Db |||||
QY 481 ITKARQNKWETGNNLDVLTYNPSYPNLDHORSVLFINKKYFLVIDRAIGATGNLGVHW 540
Db |||||
QY 541 QLKEDSNFVFKTRNVTYTRDGNLMIQSLNADRTSLNEBEGKVSYYNKELKRPV 600
Db |||||
QY 541 QLKEDSNFVFKTRNVTYTRDGNLMIQSLNADRTSLNEBEGKVSYYNKELKRPV 600
Db |||||
QY 601 FEKPKNAGTQNFVSIYVYDQKAPESIRENKGNDFEKGKLNLTITNGKQQLVLP 659
Db |||||
QY 601 FEKPKNAGTQNFVSIYVYDQKAPESIRENKGNDFEKGKLNLTITNGKQQLVLP 659
Db |||||

RESULT 2
A:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: A97909
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; H
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 193, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; PMID:21429245; PMID:11544234
A:Accession: A97909
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-633 <KUR>
A:Cross-references: GB:AE007317; PIN:AAK99101.1; PID:g15457850; GSPDB:GN00174
C:Genetics:
A:Gene: spr0297

Query Match 5.4%; Score 189; DB 2; Length 633;
Best Local Similarity 21.1%; Pred. No. 2e-05;
Matches 139; Conservative 100; Mismatches 273; Indels 148; Gaps 28;
QY 71 SKAREPDFSNAEKPADIRQIDKVTREMAKALVHOFQPHKGYGFDYDKDINQWMPVK 130
Db |||||
QY 131 DNEVWQLHRVQWQAMALVYHATGDEKYAR-----EWYQYSDWAKRNPLGLSQDND 183
Db |||||
QY 73 DPEWYMLNRQBYLLQFMIGYLVGDKGYIQCKFFLDWIEQVREFSPQSLMTLTDGT 132
Db |||||
QY 184 --KFVWRPLEVSDRVQSLPPTFSLFVNSPAFTPAFLMEFLNSVHQADYLSHYABQ--- 238
Db |||||
QY 133 IRSFTWLK-----LLLLKFDLLEKELEKILVLEKQIDFMKSYIRAKYTL 180
Db |||||
QY 239 GNHRLFEAQRNLFAGVSPEFKDSPRWRTGTSVLNTEIKQVYADGMQFELSPIYHV-- 296
Db |||||
QY 181 SNWGILQITPMLAIYHFFSDKMDLEAHYFASEELKQIETQILGDGSOFEQSLHYVEV 240
Db |||||
QY 297 --AAIDIFLKAYSARKVNLEKEFFQSYVQTVENMIMALISLSDYNTPMFGDSWITDK 354
Db |||||
QY 241 YKALLDLCL-----LLPDQDSYQELLEKMATYIQMTGLDGRITLAFGSDSTET 290
Db |||||
QY 355 NFRMAQFA-----SWARVFPAQAIKYFATDQKQKAPNLSKALSNA 397
Db |||||
QY 291 TEMPLSASVINKEDLLNGLDVKKVLLSILFLGREKVKLQEPERRAWQPK--SMIFEDS 348
Db |||||
QY 398 GFYTFRSQWMDKNATVWVLKASPPGFHAPDNGTGFELFIKGRNFTPDAGVYVYSGDEAIM 457
Db |||||
QY 349 GHVCIKD-----EHRVLFKNGPLGSAHSHSDNSFCLOQCGQPIFIDAGRYSI----- 397
Db |||||
QY 458 KLRNWRQTRI-----HSTLTDNQNVITKARQ---NKWE-----TGNNLDVLT 499
Db |||||
QY 398 --REIYERYLLKSAWSHSTCIVD-----KAPERITGSWEVEYYPHSLFCHHKEREGMH 449

QY 500 YTNPSY-----PNLD--HORSVLFINKKYFLVID--RAIGATGNLGVHQLKEDSNPVD 551
Db |||||
QY 450 YIEGAYSAEDLPVLHKKRKLMLVEDVWLLVDDIRCOGQ-----HEALTQF---ILD 499
Db |||||
QY 552 KTKNRVYTYRDG--NNLMIOSLNADRTSLNEBEGKVSYYNKELKRPV-----FEKP 604
Db |||||
QY 500 KD-----VTYQDGKINQLRLWS-----EVDFOLEDTIISPKN--ELERSKLTIKRQFFNQ 549
Db |||||
QY 605 KKN-----ACTQNFVSI---VYPDQKAPESIRENKGNDFEKGKLNLTITNGKQQLVLP 658
Db |||||
QY 550 MLDYTIHAHSEFIRHSYQTDREVENALAFENKDETDLKLLLLSIEDIRVGEKCLIV 609
Db |||||

RESULT 3
F95038
hypothetical protein SP0327 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: F95038
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; PMID:21357209; PMID:11463916
A:Accession: F95038
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-546 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74503.1; PID:g14971800; GSPDB:GN00164; TIGR:SP
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0327

Query Match 5.0%; Score 176; DB 2; Length 546;
Best Local Similarity 20.8%; Pred. No. 0.00014;
Matches 117; Conservative 82; Mismatches 231; Indels 132; Gaps 23;
QY 71 SKAREPDFSNAEKPADIRQIDKVTREMAKALVHOFQPHKGYGFDYDKDINQWMPVK 130
Db |||||
QY 26 SKDYEKVKSLERLMDNRFMDS--PDMPEPCSKIHQIQP-----WVWQVFD 72
Db |||||
QY 131 DNEVWQLHRVQWQAMALVYHATGDEKYAR-----EWYQYSDWAKRNPLGLSQDND 183
Db |||||
QY 73 DPEWYMLNRQBYLLQFMIGYLVGDKGYIQCKFFLDWIEQVREFSPQSLMTLTDGT 132
Db |||||
QY 184 --KFVWRPLEVSDRVQSLPPTFSLFVNSPAFTPAFLMEFLNSVHQADYLSHYABQ--- 238
Db |||||
QY 133 IRSFTWLK-----LLLLKFDLLEKELEKILVLEKQIDFMKSYIRAKYTL 180
Db |||||
QY 239 GNHRLFEAQRNLFAGVSPEFKDSPRWRTGTSVLNTEIKQVYADGMQFELSPIYHV-- 296
Db |||||
QY 181 SNWGILQITPMLAIYHFFSDKMDLEAHYFASEELKQIETQILGDGSOFEQSLHYVEV 240
Db |||||
QY 297 --AAIDIFLKAYSARKVNLEKEFFQSYVQTVENMIMALISLSDYNTPMFGDSWITDK 354
Db |||||
QY 241 YKALLDLCL-----LLPDQDSYQELLEKMATYIQMTGLDGRITLAFGSDSTET 290
Db |||||
QY 355 NFRMAQFA-----SWARVFPAQAIKYFATDQKQKAPNLSKALSNA 397
Db |||||
QY 291 TEMPLSASVINKEDLLNGLDVKKVLLSILFLGREKVKLQEPERRAWQPK--SMIFEDS 348
Db |||||
QY 398 GFYTFRSQWMDKNATVWVLKASPPGFHAPDNGTGFELFIKGRNFTPDAGVYVYSGDEAIM 457
Db |||||
QY 349 GHVCIKD-----EHRVLFKNGPLGSAHSHSDNSFCLOQCGQPIFIDAGRYSI----- 397
Db |||||
QY 458 KLRNWRQTRI-----HSTLTDNQNVITKARQ---NKWE-----TGNNLDVLT 499
Db |||||
QY 398 --REIYERYLLKSAWSHSTCIVD-----KAPERITGSWEVEYYPHSLFCHHKEREGMH 449
Db |||||
QY 500 YTNPSY-----PNLD--HORSVLFINKKYFLVID--RAIGATGNLGVHQLKEDSNPVD 551

```

Db 450 YIBGAWSAEPDLPYLHRRKILMLVEDVLLVDIRCOGQ-----HEVLTQF---ILD 499
QY 552 KTKNRVYTYTRDG--NNLMIQS 571
Db 500 KD-----VTYQDGKINQLRLWS 516

RESULT 4
F90559
conserved hypothetical protein MYPVU_3820 [imported] - Mycoplasma pulmonis (strain UAB CT)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: F90559
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallissou, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: F90559
A:Status: preliminary
A:Genetic code: SGC3
A:Molecule type: DNA
A:Residues: 1-1183 <KUR>
A:Cross-references: GB:AL445566; PID:gl4089796; PIDN:CACI3555.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPVU_3820
A:Genetic code: SGC3

Query Match 4.6%; Score 160; DB 2; Length 1183;
Best Local Similarity 18.7%; Pred. No. 0.0062;
Matches 159; Conservative 131; Mismatches 252; Indels 308; Gaps 44;

QY 7 KRIL---VPAVIAL-----SSGNILAQSSITR-----KDFD----- 35
Db 4 KKIASSLFLPMSLLPFLSNTSTIAQEKNIQTQFKKNSLENFDEYRKNNDEEIGBKTL 63
QY 36 -----HINLEYGGLKVKNAAGNVDDAAKALLAYREKSKAREPDFSNAEKPAD 86
Db 64 IVSKFIETKMLNLYKSDFEKINIVINDSNE-----NFSLGNNNSQ 105
QY 87 IRQPIDKVTREMAKALVHQFQPHKGYGVDYGDINQWMPVKDNEVRWQLHRVKWQA 146
Db 106 IE-----NWTNLNVKTNIDID-KLNSVKFF-- 128
QY 147 MALVYHATGDEKYAREWVYQYSDWARKNPLGLSQDNKDFWVRPLEVSDRVQSLPPTFSLF 206
Db 129 -----FNDKSSQSW-----DFIEITNIKQRIEKIKSLSQK 160
QY 207 VNSPAPTEA-----FLME-----FLNSYHQADYL-----STHYAEOGNHRLFEAQ 247
Db 161 YNDNHFSVSIKKLDFIDYKSPKNDLSFLNEFKKVDVILEIKRVNPFYKKEINSQIKIV 220
QY 248 RNLFAGVSPFEKDSR--WRQ-TGISVLNTEIKKQVYA---DGMQFELSPIYHVAIDI 301
Db 221 KD--FDKLNFEIFTSSLLFEKWNVLDDLHNEIYKIKIQLDFNKNKFEIK--YQILSDDL 277
QY 302 FLKAYASAKRVNLEKEFPQSY--VQTVENMIMALISISLDPYNT----- 343
Db 278 I-----SSDKPFLLLFRGEYFNQNIIDLGIKRVKISRSDINSQSPIKORLLIYKNNV 331
QY 344 -----PMF-----GDSWITDKNFRMAQFASWAR---VFPANQAIK 375
Db 332 YKNFDGNGKEKVPQFIHEERISNEYGGHVLNQPFRVAFISSQNNNEVIYVDID 391
QY 376 -----YFATDGQKAPN-----FLSKALSNAGVYTFRSGWD--KNA 410
Db 392 NIDNLFDFLEHQQRKKTKITIKTDNTKLTLYEIIYSKELKDSAIKARWENWDPINNL 451
QY 411 TVMVL-----KASPPGEFHQAPNGTGFELFI-----KGRNFTPD-----AGVF 448
Db 452 EHKKLVKHNENGKVPKNPNRINPHNGLEKEKIFWLEHEKLNHLSPPFKVFNKEGAF 511
QY 449 V-----YSGD-----EAIMKLRNVQRTRIHSITLTDNQNNVITKARQNKW-ETGNL 495

```

```

Db 512 VKVISIQKAYSIDLPPENSLTLKKYFWMV-----SNNDFSOEELITIKDNNNSFSKSGNFL 568
QY 496 DVLTYTN-----PSYPNLDHQSRVLFINKKYFLVIDRAIGEATGNLGVHQLK--- 543
Db 569 LELKNTNYSNVLNISGVGYNLKLHNL--IDKKILVPVEKSI---AGNLIKNYLLKNFN 623
QY 544 -EDSNPVDPKTKNRVYTYTRDGNLMIQSLNADRTSLNEE-----EGKV-----SYVYNKE- 593
Db 624 FEES--ISSLSYBEIVQNNKLLNVLNLEFNLTKNNKKNKIFDIENKMDSLSTYVDNKNK 681
QY 594 ----LKRPAFVEKPKKNAGTQNFVSIYVYDQOKAPEISIRENKGNDEKGLN---LT 646
Db 682 LLSLDEEFELKINSFN--TKNFVEIDYQD---HPNVDLIFEKASIFYKDKINKKQIT 736
QY 647 LTINGKQQLV 656
Db 737 ININ-KERII 745

RESULT 5
JC5870
poly(beta-D-mannuronate) lyase (EC 4.2.2.3) - Pseudomonas sp.
N:Alternate names: alginate lyase I
C:Species: Pseudomonas sp.
C:Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C:Accession: JC5870
R:Kraiawattanaspong, J.; Ooi, T.; Kinoshita, S.
Biochem. Biotechnol. Biochem. 61, 1853-1857, 1997
A:Title: Cloning and sequence analysis of the gene (alyII) coding for an alginate lyase
A:Reference number: JC5870; MUID:98067850; PMID:9404064
A:Accession: JC5870
A:Molecule type: DNA
A:Residues: 1-713 <KRA>
A:Cross-references: DDBJ:AB003330; NID:g2073106; PIDN:BAAL9848.1; PID:g2073107
A:Experimental source: strain OS-ALG-1
C:Comment: This enzyme catalyzes the degradation of alginate by a beta-elimination mecha
cleaved alginate oligomers.
C:Genetics:
A:Gene: alyII
C:Superfamily: Pseudomonas poly(beta-D-mannuronate) lyase
C:Keywords: carbon-oxygen lyase

Query Match 4.4%; Score 154; DB 2; Length 713;
Best Local Similarity 20.3%; Pred. No. 0.0076;
Matches 153; Conservative 92; Mismatches 234; Indels 276; Gaps 42;

QY 87 IRQPIDKVTREMAKALVHQF--QPHKGYGVDYGDINQWMPVKDNEVRWQLHRVKW 143
Db 48 LRQSYQAV-KNAADKALAQPIVVPVKDGGGYTH-----EQHKNY 88
QY 144 WQAM--ALVYHATGDEKYA-----REWYQYSDWARKNPLGLSQDN-----DKF 185
Db 89 SNMLNGVAYQISGEKKYADYKVNMLNVSQYKWP-LHPKRKSEEDGGRIFWQSLNDF 147
QY 186 VWR--PLEVSDRV-QSLPPTFSLFVNSPAPTEAFLMEF-----LNSYHQADYLSTH 234
Db 148 VMQLTYIQAYDLVYDGIPATDKRTIEEKLFPV--ILKFFTEPDYDVFNKIH----- 196
QY 235 YAEQGNHRLFEAQNLFA-GVS-----FPE--FKDSPRWQRTGISVLNTEIKKQVYA 283
Db 197 -----NHGTW-----NLAAVGITGVNLKREYEMAIKSGKDKGTGL---AQIDQLFSP 244
QY 284 DGMQFELSPIYHVAIDIFLKAYGSAKRVNLEKEFPQSYVQTVENMIMALISISLDPYNT 343
Db 245 DGYTME-GPYQRYALLPFVL--FAKAIN-NVEPSRKIFEYRDKLSKAIHTSL----- 294
QY 344 PMFGSWITDKNFRMAQFASWARVFPANQAIK----- 375
Db 295 ----QTSYTDKTF-----FPLNDAIKDKTYESVELVYGVLDAYADIKAEVDLLD 339
QY 376 -----YFATDGQKAPNLSKAL-----SNAGFYTFRSGWDKNATV 412

```

Db 340 IARQQNRVIVSDAGLKVRAADLAAGKAVPFKYQTLWIRDGKGDEGLGILRNGPNPDQQC 399

QY 413 MVLKASPPGEFHAQPD-----NGTFELFKGRNFTPDAG-----VFVSGDEAIMKL 459

Db 400 VVLKASQGWGHGHRUNLLFYDNTTEIF-----PDYGAARFLNIDTKGGGGLPSPN 452

QY 460 RNWYQTRIHSTLITLD-----NQ-NMVTIKARONKWT-----491

Db 453 NTWAKQTVAHNALVVDQTSHEKAGLPADKASPTLLYFSNQPNLKVSAKEDKAYTDVTM 512

QY 492 -----GNLIDLVTNPSYNPLDHRQSVLP-----INXKYFLVID-----RAIGE 531

Db 513 LRTSALVKEGDKELLIDVMOQAQSAKSHQYDLPEWYKQGLVNTSPVPTAKANQLTALGD 572

QY 532 ATGNLGVHWLKDSDNPVFDK-----KNRVYTYRDN-----LMIOSLNADRTSLNE 581

Db 573 KNGYQHI-W--LNAGNPLEGKGMVGLNKNRFFYTHFVSDNPLEVKLLISGANDPEMNL 629

QY 582 EEGKVSYYVYNKELKRPAPVFEKPKNAGTQNFVSVIYPYDG-----OKAPEIS-----629

Db 630 VDGK-----AFMLSSGGN-----QTFVSIETHGGTDPINETVSSALPTVSGLK 674

QY 630 -IRENKGNDFEKGKLN---LTVTIN---GKQQLVLV 658

Db 675 LIKSDAQQTIIISFKVNERTYTYQINVTXKQQLYII 709

RESULT 6

T36084

hypothetical protein SCE134.02c - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T36084

R:Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajadream, M.A. submitted to the EMBL Data Library, April 1999

A:Reference number: Z21596

A:Accession: T36084

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-561 <SAU>

A:Cross-references: EMBL:AL049661; PIDN: CAB41200.1; GSPDB: GNO00070; SCOEDB: SCE134.02c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB: SCE134.02c

Query Match 4.0%; Score 139; DB 2; Length 561;

Best Local Similarity 21.6%; Pred. No. 0.061;

Matches 76; Conservative 45; Mismatches 163; Indels 68; Gaps 12;

QY 273 LNTKIKQVYADGMQFELSPIYHVAADIFLKAYGSAKRVNLEKEFPQSYQTVTENMIMA 332

Db 186 LAAHMRAAVHRDGEWEASTYVHLFVHLAYLLALRGAR-----PERLPDDLAALGNMYS 241

QY 333 LISISLPDYNTPFGD-----SWITDKNFRMAQFASWAKRVFPANQAIKF 377

Db 242 LAGVAAPDGLPALHDGYPYRREAAETREVAALATQLFAGAPLKSVAQAERDAAALS 301

QY 378 A-TDCKQKQKAPNFLSKA---LSNAGFYFRSGWMDKNATVMVLKASPPGEFHAQPDNGTPE 433

Db 302 ADLDGWADFPLPATPAHRTFDGIAVFRSA-----GVHALLDAGPHGGHGHQKLSLY 357

QY 434 LFI-KGRNFTPDAGVFVYSGDEBAMKLRNWRQTRIHSHTLTLTDNQNMVITKARONKWTG 492

Db 358 LYADDGTAWQPDGQVPY-----AHRNLRITYASTAAHPTFRVDEREQQPCDAVLD-GDGT 412

QY 493 NNLDLVTNPSYNPLDHRQSVLFINKKY---FLVIDRAIGATGNLGVHWLKDSDNPV 549

Db 413 R-----CEQAYDGVGTATRRVV-TDRSYLLDILVLDAA-----443

QY 550 FDKTKNRVTVYRDGNLMIQSLNADRTSLNEEGKVSYYVYNKELKRPAPVF 601

Db 444 ---TEERLTAQLRPGTDLVVAQGTDRARTVWGDGSAVLTGHHTSRPQAEF 492

RESULT 7

AB1695

p60-related protein homolog lin2100 [imported] - Listeria innocua (strain Clip11262)

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AB1695

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U. Science 294, 849-852, 2001

A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mok, C.; Schlueter, T.; Simoes, N.; Tietze, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species

A:Reference number: AB1077; MUID: 21537279; PMID: 11679669

A:Accession: AB1695

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-764 <GLA>

A:Cross-references: GB:AL592022; PIDN: CAC97330.1; PID: g16414614; GSPDB: GNO0178

A:Experimental source: strain Clip11262

C:Genetics:

A:Gene: lin2100

Query Match 4.0%; Score 139; DB 2; Length 764;

Best Local Similarity 19.3%; Pred. No. 0.098;

Matches 148; Conservative 97; Mismatches 280; Indels 242; Gaps 37;

QY 7 KRIL-----VFVAVI---ALSSGNILAQSSITRKDFPHINLEYSGLKVKVKAAGNYDDAA 60

Db 2 KRILSVLCIFALVGLFSPFTIKASATNTKTFEEL-----YAEAK 43

QY 61 KAL-LAYREKSKAREPDSNA-----EKPAIRQIP-----IDKVTREMAKKA 102

Db 44 KHLGKPYSDQDQRRGPNYFDGSGVYQVYKVTGVRIPNTSAPQYSAADKNGKNGKQPG 103

QY 103 LVHQFQPHKGYGYFDYGDINQWQWPFVDNEVRWQLHRVKKWQMAALVY---HATGDEKXA 160

Db 104 DLVYFKGHVGI-YIGNGRMINAQNDGVKIDNI-----NSSYQSIFFVGVGRFFNSEKKG 157

QY 161 REWTVQYSD-----W-----ARKNPLG-----LSQDNDKFFVRPLEVSDRVQSIPPTFS 204

Db 158 SKSAYAVSDNLNRSSNNWDSVAGKVPQGAQVSDLDSDKNGCMVTVNTTKGYMLNTTN 217

QY 205 LFNVSPTAPFAP-----LMEFLNSYHQOQADVLSTH 234

Db 218 YPSDTFVLTKTYAKONINLRKATWDSVAQVKQGEKVTNMLKTNVNGWYQVTVGGKGTG 277

QY 235 YAEQGNHRLFEAQRNL--FAGVSFFPKDPSPEWRQTGISVLNTE---IKQVYADGMQPE 289

Db 278 YMLNNYLVENPLNMYTAYVGLNLSAANW-DSSISLVVPEGRAKVE---MDTN 331

QY 230 LSPYHYA-----AIDIFLKAYGSAKRVNLEKE---FPQSYQTVTE---NMI 330

Db 332 SGFWTKVYQYQGTGYIPLTDDYLSKTTVLTKTYAKDNLRKATWDSVAQVKQGEKVT 391

QY 331 MALISISLPDYNTPMFG-----DSWITDKNFRMAQF-----ASW---AR 366

Db 332 TVNLKTSVNGVQVTVYGGKGMILNDYILVEKALMKMTYYAVSSILNRSEAKWSSISQ 451

QY 367 VPPANQAIK-----YFAT--DGKQKAP---NFLSKALSNAGFYT-----FRSG 405

Db 452 VVPEGRAVKVEMDTNVGNWFKVYDNTKGYMPLNDLILSETAVLTKTYAKONLNRSEAK 511

QY 406 WDKNATVMVLKASPPGEFHAQPDNGTFFELFKGRNFTPDAGVFVYSGDEBAMKLRNWRQ 465

Db 512 WDSEVTVQVEKGEKVTVNSKTSIDGWYEV-----TYGKKKGYMILNNY---555

QY 466 TRIHSTLTLTDNQMV---ITKARONKWTG-----NNLDVLTVTNPS 504

Db 556 -LVABPDLTKTYAVNTNLRSSEKWDSSIQQVPEGAQKVKVEMNTSDGNWTKYQNTK 614

QY 505 --YPNLDHQRSVLFINKKYFLVIDRAIGATGNLGVHWLKDSDSNPFDKTKNRVTVTYR 562

Db 615 GYMPNDLNLSETAVLTKYY-----AKDNLNLRSEAKWDS-----EISQVVE 656

QY 563 DGNMLTQSLNADRTSLN-----BEEGKVSYYNKKELKRPFAVPEKP 604

Db 657 KGEKVTINS-----KTSINGWHEVTYGGKGYM-----ILSDNYLVEKP 695

RESULT 8
T44835
hypothetical protein weef [imported] - Acinetobacter lwoffii
C:Species: Acinetobacter lwoffii
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 19-May-2000
C:Accession: T44835
R:Nakar, D.; Gutnick, D.L.
submitted to the EMBL Data Library, July 1999
A:Description: Genomic organization of the wee region of Acinetobacter lwoffii RAG-1
A:Reference number: 222856
A:Accession: T44835
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-604 <NAK>
A:Cross-references: EMBL:A0243431; PIDN:CAB57203.1
A:Experimental source: strain RAG-1
C:Genetics:
A:Gene: weef
C:Superfamily: Acinetobacter lwoffii hypothetical protein weef

Query Match 3.9%; Score 137.5; DB 2; Length 604;
Best Local Similarity 20.5%; Pred. No. 0.087;
Matches 126; Conservative 76; Mismatches 243; Indels 169; Gaps 33;

QY 90 PIDKVTREMAKALVHQFQPH-----KGYGYFDY-----GKDINWQWMP-----128

Db 34 PVOKLSAQLPMGCFQSFQSAFTGMNVPVNSTLTAFGYLKYPVAGLPNFWFYSPLTHIFQ 93

QY 129 -----VKDNEVRQLHRVQWQAMALVYHATGDEKARE-----WV 164

Db 94 NTEKAWYQIPDFGNVDIGKIWEASRFDW--LIDLVLH-----ERQVRQCALVQLDLWL 147

QY 165 QYSDWARKNPGLSQDNKDFVR--ELEVSDRVQSLPPTFSLFVNSPFTF-----AFLM 218

Db 148 ---NDWCEKNPAYFGFN-----WKCQGEASIRVWHI---ITALIGLEQEQHPHDNVCAFT 197

QY 219 EFLNSVHQADYLSHYAEGNHLFEAQNLFAVGFPEFKDS---PRWRQTGISVLNT 275

Db 198 AHLKRIEPTIDYA---IAONNHTGEAAALYTGGAVLNLWKPKSYQKQWALGEXWLLN 254

QY 276 ETKQVYADGMQFELSPIYHVAIDIF-----LKAYGSAKRVNLEKEFPQSYVQ 324

Db 255 RATKLIMQDGGFSQYSINHRVMDLSYCLAETVRQKPELKFPGSQLYRQLQKATDMLYVL 314

QY 325 T-----VENM-----IMALLISLPDNTPMFGSDWITDKNFR--MAQFASNA-----RVFPA 370

Db 315 TQADGVPNLGANDGARLIPVSTQDY-----RDFRPTVQLASTLPHQHSYYPE 362

QY 371 ----NOAKYFATDGKQKAPNLSKALSACGYFTFRSGW--DKNATVMVLKASPPGEPH 424

Db 363 PPSYDESILTFQI--QKLQDNF--ELPSRQNFNMSGMSQNNFNFIATKLPKFKER 417

QY 425 -AQPNDGTFFELFKGRNFTPDAGVFVYSGDEALMKLNWYQRIHSTLTLDNQNMV--I 481

Db 418 PSQCDALHLDVFKGNLRLDGGTSYN---STVEDLNYPSGVASHTVEFHOHQMPRL 474

QY 482 TYARQNKWTGNMLDLVLTNTNSYP-----NLDHORSVLFINKKYFLVIDRAIGATG 534

Db 475 SRFLFGAWLTPKE--LITYSANEFQGYQDHWGCTHHRKI-----SLTDNAIKADN 523

QY 535 NLG-----VHMOLKED-----SNPFDKTKNRVITYYTRGNMLTOSLNADR--TSINEE 583

Db 524 IGFQOQAVLRWKLQPDHWTLENNLSNGKVERII-----LEAHPITLWRSE 570

QY 584 GKVSYYNKKELKRP 597

Db 571 GBSRYYYQSKVP 584

RESULT 9
E96554
hypothetical protein F19C24.18 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 15-Jun-2001
C:Accession: E96554
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E96554
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-560 <STO>
A:Cross-references: GB:AB005173; NID:g11094759; PIDN:AAG29692.1; GSPDB:GN00141
C:Genetics:
A:Gene: F19C24.18
A:Map position: 1
C:Superfamily: human mannosyl-oligosaccharide 1,2-alpha-mannosidase

Query Match 3.6%; Score 127; DB 2; Length 560;
Best Local Similarity 20.2%; Pred. No. 0.43;
Matches 89; Conservative 82; Mismatches 159; Indels 110; Gaps 22;

QY 7 KRIIVFAVIALSSNLIQAOSSTITRK---DFDHINLEYSGLEKVKAKVAGNVDRAKAL 63

Db 26 RRLALLFIIVFVSVMVLDNRINLAREHEVEFKLNEEVRLEQMLBELNGGVGNKPLKTL 85

QY 64 LAYREKSKAREPDFSNAEKPADIQPIDKVTREMAKALVHQFQPHKGYGYFDYGDIN 123

Db 86 -----KDAPED---PVDKQRQKVKEMIAHWSSYEKA---WQKD-- 120

QY 124 WQWMP--VKDNEVRQLHRVQWQAMALVYHATGDEKY--AREWVYOYSDWARKNPLGLSQ 180

Db 121 -ELQPRTKDGTDSFGGLGATMVDLSLTLYIMGLDEQFKAREWVASSLDFDKDYDASME 179

QY 181 -----DNKD--FYWRPLEVSDRVQSLPPTFSLFVNSPFTFPAFLMEFLNSY 224

Db 180 TTIRVVGGLLSAYDLSGDKNFLEKAKDIADR---LLPAW---NTPTGIPYNIILNRNG-- 231

QY 225 HQADYLSH---YABQGNHRLFEAQRNLFAGVSPFEFKDSPRWRQTGISVLNTEIKQV 281

Db 232 -----NAHNSWAAGGDSILADSTEQLFIALSORTGDPKYQKQKVKVI--TELKNFP 283

QY 282 YADGQFELSPIY-----HVAADIF---LKAYGSAKRVNLEKEFPQSYV 323

Db 284 PADG---LLPTIYNPDNANPSYSTTTFCAGMDSFVEYLLKVVQGNKTSVAKPYRDMWE 339

QY 324 QTVENMIVALISLIPDNTPMFGSWITDKN-----FRMAQFASWARVFPANQAIKYPFA 378

Db 340 KSMKG--LLSLVKKSTPSSFT-----YTCEKNGNLLDKMDLACFA---PGMLAL----- 385

QY 379 TDGKQKQKAPNLSKALSNA 398

Db 386 --GASGYGDEKFKELSLAG 403

RESULT 10
AC0422
hypothetical protein YP03474 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AC0422
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0422
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-672 <KUR>
A:CROSS-references: GB:AL590842; PIDN:CAC92703.1; PID:gi5981398; GSPDB:GN00175
C:Genetics:
A:Gene: YPO3474

	Query Match	3.6%;	Score 125.5;	DB 2;	Length 672;
	Best Local Similarity	18.9%;	Pred. No. 0.73;		
	Matches 114;	Conservative 81;	Mismatches 248;	Indels 159;	Gaps 24;
QY	87	IRQPIDKVTREMAKALVHQ	---FOPHKG	-----YGYF-DYXKDINQMW	-----PV 129
Db	17	WKQP-EIATLADNQVLETDLPATGIATWNNHYICPEHGVLVWDRFSPQAHRCV	75		
QY	130	KDNEVRWOLHVRKMQA	-----MALVYHATGDEKY	---AREWVYQSDWARKNP	175
Db	76	DDHLFSGEPYDGAWWRALNGLNAKACNQLGLLWLTGTEIRYDKVRDILMSY	---ARYYP	132	
QY	176	LGLSODNDKFWVRPLEVSDRVOSLPPTFSLFVNSAFIPA	-----FLMEPLNSYHQQA	228	
Db	133	-----DYEVHGIP	-----YNGFGKANAQTLCEANCLLDPALGYDFTA	170	
QY	229	DYLSHYAEQGNHRLFEAQRNLFAGVSF	-----PEFKDSPRWQRTGISV	-----272	
Db	171	DTLSQ	-----BQDCLAEHLRVGADFLMQHRTQLHNEHVKISAIIVIGLILEEHY	224	
QY	273	-----LNTEIKQVYADGMQFELSPYHYVAAIDIFLKAYGSAKRVNLEKFPQSY	322		
Db	225	IEFAVNADYGLAYQLEHGLFNEGLWFGSVHYHYA	---LQGFWSFEKLAAGSRYSLLA	280	
QY	323	VOTVENMIMALISISLPYNTPMFGDSWTDKRNFRMAQFASWARVFPAN	-----371		
Db	281	LFYRDMLSFPLKLLMNGTFRINDCTAGQQLNAHLYEFAYQIYQVREYAAALQHIY	340		
QY	372	-----QAIKYPATDGKQKQKAFNLSKALSNAGFYTFRSGMDKNAT	411		
Db	341	ROPRNLNDALLYGAESLPLOLDIVPTDLH	--AP-----DCGLTILQ--POASR	388	
QY	412	VMVLKASPPGEPHAQPDNGTGFELFIKGRNFTPDAGVYVYSGDEAIMKLRNWRQTRIHT	471		
Db	389	ALLIKHSPYGGHHDYRLNLILFDRGHEVLPDLGTTGYA	---QRHYDYKNSATHNT	444	
QY	472	LFLDNQNVITKARQNKWETGN	---LDVLTYNPSYPNLDHORSVLFINKKXFLVIDRA	528	
Db	445	LSINQKNQPPAVPYIRQWHQAANFSLDTEVDWROTPAPELDSHTRVQWDEASY	-----RD 499		
QY	529	IGBATGNLGHVWOLKEDSNPVDKTKNRVYTYRDCGNLMIMQSLNADRTSLNEEGKVS	588		
Db	500	VKERRILLWLEDTIIDISINVPHQSQSWNTLYIDGT	---ATEQKGMTSFG-DSGPMQY	555	
QY	589	VY 590			
Db	556	VH 557			

RESULT 11
GB2875
hypothesized protein U0558 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: GB2875
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mid

A:Reference number: A82870
A:Accession: G82875
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1883 <GLA>
A:CROSS-references: GB:AB002154; GB:AF222894; NID:G6899557; PIDN:AAF30971.1; GSPDB:GN000
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: U0558
A:Genetic code: SGC3

	Query Match	3.6%;	Score 124.5;	DB 2;	Length 1883;
	Best Local Similarity	20.1%;	Pred. No. 4.3;		
	Matches 118;	Conservative 74;	Mismatches 200;	Indels 195;	Gaps 31;
QY	132	VSDRVQSLPPTFSLFVNSPAPTAPFLMEFLNSYHQQADYLSHYAEQGNHRLFEAQRNL	251		
Db	631	VTDRY-----PYLSLVAQS---DPAFKTTGIN	-----YLKYVSTH---EYGHQTLQDMKOLS	677	
QY	252	-----AGVSFPPEF	-----KDSPRWQRTGISVLNTEIKKQVYADGMQFELS	291	
Db	678	DSNDSVIGGIDSRSGVSDSYNGKALQDYLNARSSGITFRKTDVYNPTKDGSGFFNFS	737		
QY	292	-----PIYHVAADIFLKAYGSAKR	-----VNLEKEPPQSIVQTVENMIMA	-----332	
Db	738	LNNDPKNPVWETQK-DIF	-----GSVNADDPKFAFFYNNKRRFLQKYDELFEAAKLNRVQPY	792	
QY	333	-LLISLSLPDYNT	---PMEGDSWITDKNFRMAQFASWARVFPANQAIKYFATD	---GKQK 385	
Db	793	DLFINMSFDHESATVNPSPGPD-INKDPSRLKAEY	-----YFYNDQNSQOK 837		
QY	386	APNF-LSKALSAGFYTF	---RSG-----WDKNATVMVLKASPPGEPHAQPDNGTPEL	434	
Db	838	NDNFKGVSIEKPGLSKYDGLKDCMGTPIKESKQDGRAIVYLKHKHKKPKKDD	---IBI 894		
QY	435	FIKGRNFTP	---DAGVYVYS-GDEAIMKLRNWRQTRIHTSLTLDNQNVITKARQN-KWE	490	
Db	895	LIKTNTNPVIDLSTCLSKSDGTINTRKLNKVRBIQ	-----DSINSILVKNYNGWD 947		
QY	491	TCNNLDV	-----LTYNPSYPNLDHORSVLFINKK	-----520	
Db	948	ENGNFDTSMFNLTVDHPMTSNEKRWAEIRITAMFKYDPFNLSKSLKDEKANASSV	1007		
QY	521	--YFLVIDRAIGEATGNLGHVWOLKEDSNPVDKTKNRVYTYRDCGNLMIMQSLNADRTS	578		
Db	1008	PYKVKLSQIGS	-----DIINKTQDKYALLTLQYNASKIWDNKSKEVSK 1055		
QY	579	LINEEGKVSYNKELKRPAP	-----VFEKPKQKAGTON 612		
Db	1056	LDPELIEIKKYDYDKQFDRFGVKNNVVAQTTLTFNYFDAGIEGNNGYKYFVKPKKELYQN	1115		
QY	613	FYSIVPYDQKAPK	---ISIR-----ENKGNDFE---KGKLN 645		
Db	1116	IDR	-----TKTESVESIIIGIRFSYLTQNKITSYEQLFNKSMLN 1155		

RESULT 12
IPY
beta-fructofuranosidase (EC 3.2.1.26) 2 precursor - yeast (Saccharomyces cerevisiae)
N:Alternate names: beta-D-fructofuranoside fructohydrolase; invertase; protein YII162
C:Species: Saccharomyces cerevisiae
C:Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 21-Jul-2000
C:Accession: A00899; S50365; S24557; S05871; S05870; S59658; A27748; A45656
R:Tausig, R.; Carlson, M.
Nucleic Acids Res. 11, 1943-1954, 1983
A:Title: Nucleotide sequence of the yeast SUC2 gene for invertase.
A:Reference number: A00899; MUID:83168934; PMID:6300785
A:Accession: A00899
A:Molecule type: DNA
A:Residues: 1-532 <TAU>
A:CROSS-references: EMBL:V01311; NID:G3833; PIDN:CAA24618.1; PID:G3834
R:Lyne, G.; Bowman, S.; Churcher, C.
submitted to the EMBL Data Library, December 1994

QY	148	ALVTHATGDEKIAEWVYQYSDMAKNPGLGSDNDKFWVRPLEVSDRVQSLPPTFSLFV	207
Db	76	---HATSD---LTSW-BDQPIAIAPKN---	97
QY	208	NSPAFTPAFLMEF-----LNSVHQADYLSHYAEQGNHRLFE	245
Db	98	DSGAFSGSMVVDYNTSGFFNDTIDPRQCVAIWYNTPESEQYIS--YSLDGGYFTTE	155
QY	246	AQRNLFAGVSPPEFKD-----SPRWROTGISVLNTEIKQVYAD-----	284
Db	156	YQKNPVLAAANGSTQPRDPKVFVWPSPQKWTM--AAKSQDYKLEIYSSDDLKSKWLESAA	213
QY	285	-----GQGFEL-----SPIYHVAIDI-----FLKAY	306
Db	214	NEGFLGYQYECGLIEVTEQDPKSKYWMFISINPGAPAGGSFNQYFVGSGFNTHFEAF	273
QY	307	GSARKVNLEKEFFQSY--VQTVENMIMALISLSLDYNTPEFGDSWITDKFERMAQFA--	362
Db	274	DNQSRV--VDFGKYALQTFN-----TDITYGSAU-GIAWAS--WNEYSAFVPT	319
QY	363	-----SWARVFPANQAIKYPFA--TDGKGKAPNFLSKALSNAAGYTFTRSGWDKNATV	412
Db	320	NPWRSSMSLVKFSLN--TEYOANPETELINLKAEPILN--LSNAG-----PWSRFATN	369
QY	413	MVLKASPPGERHAQPDNGTPELFIKGRNFTPDAGVVFVSGDEATMK-----LRNRYQTR	467
Db	370	TTLTKANSYNDLNSGTGLEFEL-----VYAVNTTQTISKVSFADLSWFKGLE	419
QY	468	-----IHSTLTLDNQNMVITKARONKWETG--NNLDVLVTNPSYEN---507	
Db	420	DPEYLRMGFEVVSASSFFLDGRGNSKVFEVKNPFYTRNMSVNNQPFKSENDLSYVKYGL	479
QY	508	LDHORSVLFIN-----KKFLVIDRAIGCATGNLGV	538
Db	480	LDQNILEYFDGWDVSWNTYFMTTGNALGSVNMVTGV	517

RESULT 13

H90533

Hypothetical protein MYPV 1760 [imported] - Mycoplasma pulmonis (strain UAB CT13)

C/Species: Mycoplasma pulmonis

C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001

C/Accession: H90533

R/Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; M

Nucleic Acids Res. 29, 2145-2153, 2001

A/Title: The complete genome sequence of the murine respiratory pathogen Mycop

A/Reference number: A99512; MUID:21267165; PMID:11353084

A/Accession: H90533

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-625 <KUR>

A/Cross-references: GB:AL445566; PID:gl4089589; PIDN:CAC13349.1; GSPDB:GN00153

A/Experimental source: strain UAB CT13

C/Genetics:

A/Gene: MYPV 1760

A/Genetic code: SGC3

Query Match	3.4%;	Score 120.5;	DB 2;	Length 625;
Best Local Similarity	20.9%;	Pred. No. 1.5;		
Matches 103;	Conservative 61;	Mismatches 193;	Indels 137;	Gaps 26;

QY	225	HQADYLSHYAEQGNHRLFEAORNLFAGVSEFPEFKDSPRWROTGISVLNTEIKQVYA	283
Db	74	HGQIDGVQDFFFTQ--EELEKVDKEIFEKLSFGPEI-----FAIRKTSFNDTEV-----	120
QY	284	DGMQFELSPIYHVAIDIIFLKAYGSARKVNLEKEFFQSYQTVENMIMALISLSLDY--	341
Db	121	-----LDPKAKGQYIPITREIIVS--VTSEADLNSF---EQLKLAIMPFIHFYMH	167
QY	342	---NTEPMFGDSWITDKNFRMA-----QFASWARVF--PANQAIKFATDGKQGA---	386
Db	168	HFSNEYLYKSGSEVNDKTLQDKTGQNEILTLNRRKFVETFOALRY--TDNCONRTIIS	225

QY 387 -----PNFLSKALSNAGFYTFPSGWDK---NATV---MVLKASPPGFBHQAOPDNGTPEL 434
DB 226 QKSNPNFIGNYASDIKYKANSQDTRGLNSAIDEIPLFKIDP--RTHSHFWKFNFD- 282
QY 435 FIKGRNFTPDAGVYVSGDEALMKLNWYRQTRIHSHTLTDNQNMVITKARQKWETGNN 494
DB 283 ---NFTKEKIQYYSQELVPR-----EFQKIAPLPYNSQKT----- 318
QY 495 LDVLYTTPSYNLDHQRSLVFIKKYFLVIDRATGEATGN-LGVHWQLKE--DSNPVFD 551
DB 319 -----TNPELNKEDDE-----FSQDGFGLVRLKLTSSVTPNAPAFDWSSTAILNGPTLH 367
QY 552 KTKN-----RVVTYRDG---NNLMIQSLNADRTSLNEE 582
DB 368 KNENVVLGDNFFLNPYGGNIYVFKDIYGVSPRTYNEKAFMKNVTWNEFYNAFLEAMN-Y 426
QY 583 EGVSVYVY-----NKLKRPAPVFEKPKKNACTQNFVSVIYPYDGGQKAPESIRENKGND 637
DB 427 KSEISQLFVKNESNLKSKAKV--NPGKNHVPRTGFLTKKAKKAYKALVLTDEKGN 484
QY 638 FEKGLNLTITING 651
DB 485 KRQSNLKP-LTRNG 497

RESULT 14

E70533
probable sulfatase (EC 3.1.6.-) atsb - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: E70533
R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skeiton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sultston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: E70533
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-970 <COL>
A: Cross-references: GB:Z96070; GB:AL123456; NID:G3261791; PIDN: CAB09444.1; PID: e320915;
A: Experimental source: strain H37Rv
C: Geneticks:
A: Gene: atsb
C: Keywords: sulfuric ester hydrolase
F: 258/Modified site: 3-oxoalanine (Cys) #status predicted

Query Match 3.4%; Score 120.5; DB 2; Length 970;
Best Local Similarity 20.0%; Pred. No. 2.9;
Matches 85; Conservative 60; Mismatches 150; Indels 129; Gaps 22;
QY 131 DNEVRQLHRVKKWQAMA--LVYHATGDEKYAREWYQVSDWARKNPLGLSQNDKFEVWR 188
DB 386 DKALEW-LHTVRAQNAKTPWMLYATG-AUTHAPHHVFK--EWADKYRGEDDGDWVVRQK 441
QY 189 PLEVSBRVQSLPTFSLFVNSPAPTFAPLMEFLNSVHQADYLSHVAEQGNHRLFEAOR 248
DB 442 TFERQRLGIIPDPAEL-TERPOLFPAP-----DSMS-----EAQRLLFARQM 483
QY 249 NLFAGVS-----FPFEKDSPR-----WRQTGISVLNT---EIKKQVYADGM 286
DB 484 EVFAGSEADWNVGRLLDALEDIGESDNTLVFYINGDASMEGINTGSENFMTFLNGL 543
QY 287 -----QFELSPIY-HVAAI-----DIFLKAYGSAKVNLE--KEPPQSYVQVTENMIM 331
DB 544 DLDAERQLELTIEQYGGIAALGDEFTAPHFASAHASNTPLQWKGQWASHLGGTRDPLV 603
QY 332 A-----LISLSLP-----YNTPMFGDSWITDKNPRMA 359
DB 604 ANPARIIPDGRVRSQFTHCIDIAPTVLAIGLPPTHVDGFEQEPMDGTSEV----- 655

QY 360 QPASWARVFPANQA-----IKFYATDGGKQKAPNFLSKALSNAAGFYTF-----RSGWDKWA 410
DB 656 -----RIFDDAEDRHTVQYF-----ENFGSRAIYKDGWACARLQKAPWDLSP 700
QY 411 TVMVLKASPPGFBHQAOPDNGTPELFIKGRNFTPDAGVYVSGDEALMKLNWYRQTRIHS 470
DB 701 ETM--RRFAPGTY--DPDQDVWELYPDDFSAQKLAEAHPDKVAELTQLWQBAERNR 756
QY 471 TLTL 474
DB 757 VLPL 760

RESULT 15

E70593
hypothetical protein MYPV_6510 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: C90593
R: Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.
Nucleic Acids Res. 29, 2145-2153, 2001
A: Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pu-
A: Reference number: A9512; MUID: 21267165; PMID: 11353084
A: Accession: C90593
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-1272 <KUR>
A: Cross-references: GB:AL445566; PID: G14090066; PIDN: CAC13824.1; GSPDB: GN00153
A: Experimental source: strain UAB CTIP
C: Geneticks:
A: Gene: MYPV_6510
A: Genetic code: SGC3

Query Match 3.4%; Score 120; DB 2; Length 1272;
Best Local Similarity 20.1%; Pred. No. 4.9;
Matches 139; Conservative 96; Mismatches 218; Indels 238; Gaps 40;
QY 77 DFSN-AEKPADIRQPIDKVTREMAKALVHQFQPHRGYGYDYDGK-DINQWMPVKDNEV 134
DB 558 DFSNKKDLASSVWMDKIIE-----SLVVK---NKTENFDNFKYDISVSVSSL-DEV 608
QY 135 RQOLHRVKKWQAMALVYHATGDEKYAREWYQVSDWARKNPLGLSQNDKFEVWRPLEVSD 194
DB 609 NGKL-KIK-----MTIFKTKDR--LKEFLYEVVGF-----QLGITKN----- 644
QY 195 RVQSLPTFSLFVNSPAPTFAPLMEFLNSVHQADYLSHVAEQGNHRLFEAQ-----RNL 250
DB 645 -----LISYRNDKLSADQIIREFYVLGNATPEQQLKQLSL 683
QY 251 FAGVSPE-----FKDSPRWRTGI-----SVLNTETKKQVYADGMQF 288
DB 684 FS-VAIPDGLPEPEFISFKAKANSNDRGILTYTLRKNKTNARSGKINVEKTHNIIDSYLA 742
QY 289 ELSPIYHVAAI-DIFLKAYGSAKVNLEKEPPQSYVQVTENMIM-----ALIS----- 335
DB 743 Q-----HVLVNEIALENDSPFKNSINELSPN--QILENITLFDKNKSIITPEKDIOV 794
QY 336 -----ISLPDYNTPM--FGDSWITDKNFRMAQF-----ASWARVFPANQAIKYFA 378
DB 795 KYKIANPQYNLKQNSINVEIFEKNGHSQKVTRELVNGFKPKFDASWFDVVAKNLIKDNP 854
QY 379 TDGKQKQKAP-NFLSKALSNAAGFYTFRSQWMDKNAITMVL-----KASPPGFBH 425
DB 855 RDIKQDKTHPKFIANTFGTKSFY-FK--FQNNQTSKELLKHVNSVDNIEYKNVAGEIH- 910
QY 426 QPONGTPEL-FIKGRNFTPDAGVF---VYSGDEALMKLNWYR---QTRIHSHTLTDN- 476
DB 911 -----FDVKFTKSSSSSTETNIVIRKVSFGKDKLIFTDFEQYDDEDEHRSIALTK 964
QY 477 --QNMVITK---ARQNKWETGNNLDVLT--YTNPSYFNLDRHSVLFIKKYFLVIDRAI 529
DB 965 MPRNQFIQKIVEARQN-----NDLDALVDYFTNLIAPFRISNETL-----SFSIRQNM 1013

```
QY 530 GEATGNLGVHWOLKED-----SNPVEDKTKN----- 555
Db 1014 GSPRKNKGAAH---REDIADQTAEQFTILHRLSNLVTNKEKNIRDHIREWLNETVFAIRTS 1070
QY 556 RVYTTYRDGNNLMIQSLNADRTSLNEE-----EGKSVYVYNKELKRPAPVFEKPKKNA 608
Db 1071 HLKTSSEPKNEFIAKTSSERIELIKKYDLYIPEGYTLNIYN----- 1113
QY 609 GTQNFVSVIYPYDGGOKAPEISIRENKGNDPE 639
Db 1114 -THNITSLKYKYNQVE-----KENKLSDFE 1138
```

Search completed: July 28, 2003, 20:53:09
Job time : 58 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2003, 10:23:59 ; Search time 393 Seconds
(without alignments)
3459.332 Million cell updates/sec

Title: US-09-802-285a-2

Perfect score: 3494

Sequence: 1 MTTKFKRIIVFAVIALSG.....KGKMLTLTINGKQQLVLP 659

Scoring table:

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DBV=xlh
-Q=/cgn2_1/USPTO.spool/US09802285/runat_28072003_190514_9775/app_query.fasta_1.839
-DB=Published Applications NA -OEMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09802285 @CIGN 1.1 221 @runat_28072003_190514_9775
-NCPU=6 -TCPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELETE=7

Database : Published Applications NA:*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq1:*
11:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
12:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq3:*
13:	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
14:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
15:	/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
16:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	166.5	4.8	11427	11	US-09-070-927A-165	Sequence 165, App

C	2	116	3.3	4702	8	US-08-781-986A-268	Sequence 268, App
	3	116	3.3	14121	12	US-09-802-640-31	Sequence 31, Appl
	4	116	3.3	14121	12	US-09-920-033-3	Sequence 3, Appl
	5	115.5	3.3	3996	11	US-09-801-368-369	Sequence 369, App
	6	115	3.3	3222	11	US-09-974-300-2654	Sequence 2654, App
C	7	115	3.3	1830121	15	US-10-329-960-1	Sequence 1, Appl
	8	114.5	3.3	1830121	15	US-10-329-960-1	Sequence 1, Appl
	9	113.5	3.2	4765	15	US-10-090-624-5	Sequence 5, Appl
	10	112.5	3.2	6720	11	US-09-070-927A-321	Sequence 321, App
	11	111	3.2	2253	11	US-09-993-292A-19	Sequence 19, Appl
	12	111	3.2	8908	11	US-09-993-292A-18	Sequence 18, Appl
	13	110.5	3.2	1689	10	US-09-815-242-4610	Sequence 4610, Ap
	14	110.5	3.2	1737	10	US-09-815-242-8108	Sequence 8108, Ap
C	15	110.5	3.2	10194	11	US-09-070-927A-91	Sequence 91, Appl
	16	110	3.1	14070	11	US-09-870-759-127	Sequence 127, App
	17	110	3.1	14635	11	US-09-880-107-2287	Sequence 2287, Ap
	18	109	3.1	32480	11	US-09-847-101B-23	Sequence 23, Appl
	19	109	3.1	34427	12	US-09-111-911-5	Sequence 5, Appl
	20	109	3.1	35408	15	US-10-155-649-3	Sequence 3, Appl
	21	109	3.1	35871	10	US-09-956-335-2	Sequence 2, Appl
	22	109	3.1	35935	10	US-09-725-720-43	Sequence 43, Appl
	23	109	3.1	35935	11	US-09-782-378A-4	Sequence 4, Appl
	24	109	3.1	35935	11	US-09-782-378A-5	Sequence 5, Appl
	25	109	3.1	35978	10	US-09-956-335-1	Sequence 1, Appl
	26	109	3.1	36620	12	US-09-952-060-30	Sequence 30, Appl
	27	109	3.1	37474	12	US-09-952-060-25	Sequence 25, Appl
	28	109	3.1	38519	12	US-09-952-060-28	Sequence 28, Appl
	29	108	3.1	4938	15	US-10-055-794-1	Sequence 1, Appl
	30	108	3.1	6525	15	US-10-055-794-3	Sequence 3, Appl
	31	107.5	3.1	4011	12	US-09-934-455-21	Sequence 21, Appl
	32	107.5	3.1	4011	15	US-10-278-173-127	Sequence 127, App
C	33	107	3.1	5926	11	US-09-070-927A-220	Sequence 220, App
	34	106	3.0	1703	11	US-09-974-300-644	Sequence 644, App
	35	106	3.0	4896	10	US-09-740-274-3	Sequence 3, Appl
C	36	106	3.0	10317	10	US-09-849-866-1	Sequence 1, Appl
	37	105.5	3.0	3504	10	US-09-815-242-7425	Sequence 7425, Ap
	38	105.5	3.0	19031	11	US-09-070-927A-167	Sequence 167, App
	39	105	3.0	2979	12	US-09-842-484A-1	Sequence 1, Appl
	40	105	3.0	2979	15	US-10-184-485-4	Sequence 4, Appl
	41	105	3.0	3540	12	US-09-759-130B-41	Sequence 41, Appl
	42	104	3.0	5547	15	US-10-242-056-48	Sequence 48, Appl
	43	104	3.0	7551	10	US-09-817-514A-1	Sequence 1, Appl
	44	104	3.0	7551	15	US-10-242-056-46	Sequence 46, Appl
	45	103.5	3.0	3827	11	US-09-870-759-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1

US-09-070-927A-165/c

; Sequence 165, Application US/09070927A

; Patent No. US20020120116A1

; GENERAL INFORMATION:

; APPLICANT: Charles A. Kursch

; Patrick J. Dillon

; Steven Barash

; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 982

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/070,927A

; FILING DATE: 04-May-2000

CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/046,655
 FILING DATE: 1997-05-16
 APPLICATION NUMBER: 60/044,031
 FILING DATE: 1997-05-06
 APPLICATION NUMBER: 60/066,009
 FILING DATE: 1997-11-14
 ATTORNEY/AGENT INFORMATION:
 NAME: Kenley K. Hoover
 REGISTRATION NUMBER: 40,302
 REFERENCE/DOCKET NUMBER: PB369
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 165:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 11427 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 165:
 US-09-070-927A-165

Alignment Scores:
 Pred. No.: 3,31e-07 Length: 11427
 Score: 166.50 Matches: 121
 Percent Similarity: 35.18% Conservative: 89
 Best Local Similarity: 20.27% Mismatches: 215
 Query Match: 4.77% Indels: 174
 DB: 11 Gaps: 26

US-09-802-285A-2 (1-659) x US-09-070-927A-165 (1-11427)

QY 120 LysAspIleAsnTyrGlnMetProValLysAspAsnGluValArgTyrGlnLeuHis 139
 Db 7871 AAAGATGCGTGAATCGGTATCTGATGATGATCCAGATGGCTCTTATGTGAGT 7812
 QY 140 ArgValLysTyrTyrGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyr 159
 Db 7811 CQTCAAGCTTCTTGTAGATCGGCACAGCATATGCATTTACTTAAAGAAAGCGTTAC 7752
 QY 160 AlaArgGluTyrValTyrGlnTyrSerAspTyrAlaArgLysAsnProLeuGlyLeuSer 179
 Db 7751 TTACAGAAATGCACAGCTTGCTTATGATTTATT-----AACGATGAGGTGAGCCA 7698
 QY 180 GlnAspAsnAspLysPheValTyrArgProLeuGluValSerAspArgValGlnSerLeu 199
 Db 7697 AATTCGACGAATAGGATGTTGGCTCGGTAGATGTTGGATTCGAGTAACAAC--- 7641
 QY 200 ProProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGlu 219
 Db 7640 -----TGATGAAAAGCTTGACGTATATTTCCAAATCGCTGATTCAGA 7599
 QY 220 PheLeu-----AsnSerTyrHisGlnGlnAlaAspTyrLeuSer 232
 Db 7598 CTATTAGGAATGTAGTGTGTGAACAACGCTTGTGATCCATCTGGACTATTGGAG 7539
 QY 233 ThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGlu----- 245
 Db 7538 CGGTCTTATATCGATAA-----TACAGGTCTAGTAATTTGGGTGTGTTGGCAATTTGGT 7485
 QY 246 -----AlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAspSerPro 263
 Db 7484 GGAATGGCAGCTATTGATTTATTT-----CTTCAGAACTGGTGACCACT--- 7440
 QY 264 ArgTyrArgGlnThrGlyLysValLeuAsnThrGluIleLysLysGlnValTyrAla 283
 Db 7439 AAACAGAGGATCTAATATGGTCTCTGTTACTGAACAGCTGTGATCTACAAATTTCTATCA 7380
 QY 284 AspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeu 303
 Db 7379 GATGGAAATTCATTGGGACGACGACCGCTGTAC----- 7347

QY 304 LysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrVal 323
 Db 7346 -----AGC-----CAGCAGAAAGTTTGTGATGACATTCGTG 7320
 QY 324 GlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuPro----- 339
 Db 7319 TATCTATTGCAG-----ATTCTGAATATCTTGAAGTCAGTACCATTAGATCTTCGC 7266
 QY 340 ---AspTyrAsnThrProMetPheGlyAspSerTyrIleThrAspLys----- 354
 Db 7265 ATGAAATTAACAAACACCTATTTCTACCCATTATTTGGCGATAACCAAGATATCTTA 7206
 QY 355 -----AsnPhe----- 356
 Db 7205 AATCCGATCAATGATAGTATGATCATGTCACTTTTCATTACGTATATATATATATCGCAAA 7146
 QY 357 -----ArgMetAlaGlnPheAlaSerTyrAla 365
 Db 7145 TTAGGTTTCATATTTGAACCTCCATGACTGCGAATATGGCAAGGCTT-----TGCAG 7092
 QY 366 -----ArgValPheProAlaAsnGlnAlaIleLysTyrPheAla--- 378
 Db 7091 GGGGATCTTTATGAAGAAAGATCTGGAAAC-AATGAAGCCAAAGAACTTTTCGTGG 7033
 QY 379 -----ThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeu 394
 Db 7032 CGAATCAAGTGGCTGATGGCTACAAAGCAG-----GGATATCTA----- 6991
 QY 395 SerAsnAlaGlyPheTyrThrPhe---ArgSerGlyTyrAspLysAsnAlaThrValMet 413
 Db 6990 -----TTTACACTTTTACGGTCTGCATG----- 6964
 QY 414 ValLeuLysAlaSerProProGlyGlu-PheHisAlaGlnProAspAsnGlyThrPheGlu 433
 Db 6963 -----GAGCGCATGCTCATGCTTCTACAGGTGGATTTAC 6928
 QY 433 uLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAs 453
 Db 6927 ACTACAACTACAGGGGATGACTTATTTCCGATAGTGTGCTGTACAGTATGTCAACAA 6868
 QY 453 pGluAlaIleMetLysLeuArgAsnTyrTyrArgGlnThrArgIleHisSerThrLeuTh 473
 Db 6867 ATCAGAGCGACTTCAGCTA-----AAAGAGTGGCTTCGCACAATACCATGTT 6820
 QY 473 rLeu---AspAsnGlnAsnMetValIleThr-----LysAlaArgGlu 486
 Db 6819 TATCGCAAAAATCCCATACCTTTAGTTTCCGATACGTGGGTTATGACAAATTACCGAC 6760
 QY 486 nAsnLysTyrGluThrGlyAsnLeuAspValLeuThrTyrThrAsnProSerTyrPr 506
 Db 6759 ACCCTTATTCAGCAATAAAGAACTTTCTGTCGTTTTTTTGCAGATGCGGCTGGCT 6700
 QY 506 oAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLysTyrPheLeu----- 523
 Db 6699 GGATAAGCGGATCAGAAATCCAATGATTTTGGCGGAGCTTCATCTATTTAAAGTCGAT 6640
 QY 524 -----ValIleAspArgAlaIleGlyGlu-----AlaThr 533
 Db 6639 CAATCCGTAGTTATTATGATGCTTTGACAGCAGCAAGAAAGAGACTGAAATTACGATAC 6580
 QY 533 rGlyAsnLeuGlyValHisTyrGlnLeuLysGluAspSerAsnProValPheAspLysTh 553
 Db 6579 CTATAATTTGGCACCCTCGATAAATTTGTCAAAAGAGAGCGCATCGGTTTGTCTTAACATC 6520
 QY 553 rLysAsnArgValTyrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAs 573
 Db 6519 GAACAAGCATAAAGTACACACTCTCTTTTGTGCTGAGGC----- 6483
 QY 573 nAlaAspArgThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsnLysGlu 593
 Db 6482 -----CAACACAGCAATCAGTGGGCAAAAGGTTTCAGAGATTTTAAACCACT 6436


```
Db 1930 -----AAATGGTTCCAAACA-----TACAAAT 1910
Qy 392 LysAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTyrAspLysAsnAlaThr 411
Db 1909 AAATCATTTAGTGGCGAATTCAGATAGCTTCAGAGTCAACAGACATTAATGTTTCG 1850
Qy 412 ValMetValLeuLysAlaSerProProGlyGluPheHisAlaGlnProAspAsnGlyThr 431
Db 1849 ACTTTAAATTGAAAA-----CAAAATTCATTA 1823
Qy 432 PheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSer 451
Db 1822 TTGTGAC-----GAGCAAAATACAGCGATGATGATAAAATGTTACAAATATATAA 1775
Qy 452 GlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThr 471
Db 1774 TCGCAAAAAGATAGCGTGAACCTTGATACTAT-----ATCAATGCATTA 1730
Qy 472 LeuThrLeuAspAsnGln-----AsnMetValIleThrLysAlaArgGln 486
Db 1729 AAACAGATGGACAGCAAAATTGATCAACAATCAAGTATGCAAGATACAGGTAAAGAGAA 1670
Qy 487 AsnLysTrpGluThrGlyAsnAsnLeuAspValLeu-----ThrTyrThr 501
Db 1669 TATTAACAACACTGTTAAAGAAAACCTTAGATAAATTAAGAGAAATCATTCATCAACAG 1610
Qy 502 AsnPro-----SerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLys 520
Db 1609 TCACCATTTTCAAAGGATGATGTAAGANTATCGTAAGCAATTAACAGAAATCATCTGCA 1550
Qy 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeu----- 536
Db 1549 GATGAGCTTGCAAAATACAAAAGACTTACAAGATGCGCTAAATAGCATTAATAAGAACAA 1490
Qy 537 -----GlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAsp 551
Db 1489 GCTCAATTCGCTGAAAACCTTAGAGAAACAACTTCATGATGATATGTCACAAAGAACCT 1430
Qy 552 LysThrLysAsnArgValTyrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSer 571
Db 1429 ACAGATACAAACATTTATCTAT-----AACATGCTTAACAAAC 1391
Qy 572 LeuAsnAlaAspArgThrSerLeuAsnGluGluGluGlyLysValSerTyrValTyrAsn 591
Db 1390 TTTATAGCT-----GCAGGTTTAAATGAGGATGAAGCTAATAAATACCAAGCAATGTC 1337
Qy 592 LysGluLeuLysArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGln 611
Db 1336 AAAGACCAAAACGTTATAAAATGAATATTAATTTGAAAAACCGTTAGCAGAACACAT 1277
Qy 612 AsnPheValSerIleValTyrProTyrAspGlyGlnLysAlaProGluIleSerIleArg 631
Db 1276 AATTTAACAGAT-----TACGATAACCAAGTTGCGCAAGACACAAGTAGTTTG 1229
Qy 632 GluAsnLysGlyAsnAspPheGluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGly 651
Db 1228 ATTAATGATGTGTCAAAGTCAACGACTACTGAA-----ACGATTAAGTAATGAT 1178
Qy 652 LysGlnGlnLeuValLeu 657
Db 1177 ATTAATCAATTAACCTGTT 1160
```

RESULT 3

```
US-09-802-640-31
; Sequence 31, Application US/09802640
; Publication No. US20030036057A1
; GENERAL INFORMATION:
; APPLICANT: Braun, Andreas
; APPLICANT: Bonsal Aruna
; APPLICANT: Kleyn Patrick
; TITLE OF INVENTION: GENES AND POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE AND THEIR USE
; FILE REFERENCE: 24736-2048
```

```
; CURRENT APPLICATION NUMBER: US/09/802,640
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FASTSEQ For Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 14121
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (129)...(13820)
; OTHER INFORMATION: Nucleotide sequence encoding apolipoprotein B
; OTHER INFORMATION: (APOB)
US-09-802-640-31
```

Alignment Scores:

Pred. No.:	0.236	Length:	14121
Score:	116.00	Matches:	104
Percent Similarity:	30.9%	Conservative:	94
Best Local Similarity:	16.28%	Mismatches:	197
Query Match:	3.32%	Indels:	244
DB:	12	Gaps:	26

US-09-802-285A-2 (1-659) x US-09-802-640-31 (1-14121)

```
Qy 218 MetGluPheLeuAsnSerTyr-----HisGlnGlnAlaAspTyr 230
Db 9330 ATAGACTTCTCGAATACTATGCACTGTTCTGAGTCCAGCGCCAGCAAGCAAGTTGG 9389
Qy 231 -----LeuSerThrHisTyrAlaGluGlnGlyAsn 240
Db 9390 CAAGTAAGTGTAGTTCATCAGTATAAGTACACCAAAATTTCTCTGCTGGAACAC 9449
Qy 241 HisArgLeuPheGluAla-----GlnArgAsnLeu----- 250
Db 9450 GAGAACATTTATGAGGCCCTTAGGAATAAATGAGAGCAAAATCTGGATTTCTTAAAC 9509
Qy 251 -----PheAlaGlyValSerPheProGluPhe 259
Db 9510 ATTCTTTAACAAATCTCGAAATCGTCTACCTTACACATAATCACAACTCTCCACTG 9569
Qy 260 LysAspSerProArgTrpArgGlnThrGly----- 269
Db 9570 AAAGATTTCTCTATGCGAAAAAAGAGGCTTTGAAGGAATTTCTTGAAAAACGCAAGCAA 9629
Qy 269 ----- 269
Db 9630 TCATTGATTTAAGTGTAAAGCTCAGTATAAGAAAAACAAACACAGGCATTCCATCACA 9689
Qy 270 -----IleSerValLeuAsnThrGluIleLysLysGlnValTyrAlaAspGlyMetGln 287
Db 9690 AATCCTTTGGCTGTGCTTTGTGAGTTTATCAGTCAGAGCATCAAAATCCTTTGACAGGAT 9749
Qy 288 PheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGly 307
Db 9750 TTTCAAAAAAAC-----AGAAACAATGCATTAGATTTTGTCAACCAATCTCTATAAT 9800
Qy 308 SerAlaLys-----ArgValAsnLeuGluLys-----GluPheProGln 320
Db 9801 GAAACAAAAATTAAGTTTGTAGTACAAAGCTGAAAAATCTCACGACGAGCTCCCGAGG 9860
Qy 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
Db 9861 ACCTTT-----CAAAATTCCTGGA 9878
Qy 341 TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGln 360
Db 9879 TACACTGTTCCAGTTGTC-----AATGTTGAAGTGTCTCCA 9914
Qy 361 Phe-----AlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyr 376
Db 9915 TTACCATAGATGTCGGCATTCGGTATGTGTCTCCAAAGCAGTCAGCATGCCTAGT 9974
```



```
QY 377 PheAlaThrAspGlyLysGlnGlyLysAlaProAsnPhe----- 389
Db 9975 TTCTCCATCTCTAGGTTCTGAGTCCTGCGCTTCATACACATAATATCGCCATCATTA 10034
QY 389 ----- 389
Db 10035 GAGTCGCCAGTCCTTCATGTCCTAGAAATCTCAAGCTTCTCTTCCACATTTCAAGGA 10094
QY 390 -----LeuSerLysAlaLeuSerAsnAlaGlyPhe-----TyrThr 401
Db 10095 TTGTGTACCATAAGCCATATTTTATTCCTGCCATGGCAATATTACCTATGATTTCTCC 10154
QY 402 PheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProGly 421
Db 10155 TTTAAATCAAGT-----GTCAACACTGAATACCAATGCT 10190
QY 422 GluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPhe 441
Db 10191 GAATCTTTTAACCAAGTCAGAT-----ATTGTTGCTCATCTCTTTCTTCATCTTCATCT 10244
QY 442 ThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIle----- 456
Db 10245 GTCAATTGATGCACTGCAAGTAAATAGAGGGCACCACCAAGATTGACAGAAAAAGGGGA 10304
QY 457 -----MetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSer 470
Db 10305 TTGAAGTTAGCCACAGCTCTGCTCTGAGCAACAAATTTGTGAGGGGTAGTCATAACAGT 10364
QY 471 ThrLeuThrLeuAspAsnGlnMetValIleThrLysAlaArgGlnAsnLysTrpGlu 490
Db 10365 ACTGTGAGCTTAAACCCAGAAAAATATGGAAGTGTCAAGTGCAAAAAACCAAAAGCCGAA 10424
QY 491 -----ThrGlyAsnAsnLeuAspValLeu 498
Db 10425 ATTCCAATTTTGAGATGAATTTCAAGCAAGAACTTAATGGAATAATCAAGTCAAAACCT 10484
QY 499 ThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsn 518
Db 10485 ACTGTCTCTTCCCTCATGGAATTTAGTATGATTTCAATTTCTCAATGCTGTACTCTAAC 10544
QY 519 LysLys-----TyrPheLeuVal 524
Db 10545 GCTAAAGGAGCAGTTGACCACACAGTTAGCTTGGAAGCCCTCACCTCTTACTTTCCATT 10604
QY 525 IleAspArgAlaIleGlyGluAlaThrGlyAsnLeu----- 536
Db 10605 GAGTCATCTACCAAGAGAGATGTCAAGGGTTCGGTTCTTCTCGGGAATATTCAAGAACT 10664
QY 536 ----- 536
Db 10665 ATTGTAGTAGGCGCAACACTTACTTGAATTCACAGAGCACAGGTCTTCAGTGAAGCTG 10724
QY 537 ---GlyValHis-----TrpGlnLeuLysGluAspSerAsnProValPhe 550
Db 10725 CAGGGCACTTCCAAATTTGATGATATCTGGAACCTTCAAGTAAAAAGAAAATTTTGCTGGA 10784
QY 551 AspLysThrLysAsnArgValTyrThrThrTyrArgAspGly-----AsnAsnLeuMet 568
Db 10785 GAAGCCCACTCCACGCAATATATCTCTCTGGAGCAGCAGTACGAAAAAACCACTTACAG 10844
QY 569 IleGlnSerLeu-----AsnAlaAspArgThrSer-----LeuAsnGluGlu 582
Db 10845 CTAGAGGGCTCTTTTTCACCAACGGAGACATACAGCAAGAACCCCTGGAACTCTCT 10904
QY 583 GluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheValPheGlu 602
Db 10905 CCATGGCAATGTGTCAGCTCTTGTTCAGGTCCATCAAGTCAGCCAGTCTCTTCCATGAT 10964
QY 603 LysPro-----LysLysAsnAlaGlyThrGlnAsn----- 612
Db 10965 TTCCTGACCTGTGCCAGGAAGTGGCCCTGAATCTAACACTAGAACCAAGAGATCAGA 11024
QY 613 PheValSerIleValTyrProTyrAspGlyGlnLysAlaProGluIleSerIleArgGlu 632
```

```
Db 11025 TGAATAAATGAAGTCGGATTTCATTCTGGTCTTTCCAGGCCAGGTCGAGCTT----- 11078
QY 633 AsnLysGlyAsnAspPheGluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGly 651
Db 11079 -----TCCAATGACCAAGAAAAGGCACACCTTGACATTGCAGGATCTCTTAGAAGGA 11129

RESULT 4
US-09-920-033-3
; Sequence 3, Application US/09920033
; Publication No. US20030087853A1
; GENERAL INFORMATION:
; APPLICANT: Rosanne M. Crooke
; APPLICANT: Mark J. Graham
; TITLE OF INVENTION: ANTISENSE MODULATION OF APOLIPOPROTEIN B EXPRESSION
; FILE REFERENCE: ISPH-0592
; CURRENT APPLICATION NUMBER: US/09/920,033
; CURRENT FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 123
; SEQ ID NO 3
; LENGTH: 14121
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (129)...(13820)
US-09-920-033-3

Alignment Scores:
Pred. No.: 0.236 Length: 14121
Score: 116.00 Matches: 104
Percent Similarity: 30.99% Conservative: 94
Best Local Similarity: 16.28% Mismatches: 197
Query Match: 3.32% Indels: 244
DB: 12 Gaps: 26
```

US-09-802-285A-2 (1-659) x US-09-920-033-3 (1-14121)

```
QY 218 MetGluPheLeuAsnSerTyr-----HisGlnGlnAlaAspTyr 230
Db 9330 ATAGACTTCTGTAATACTATGCACTGTTTCTGAGTCCCAAGTCCCAAGCAAGTGG 9389
QY 231 -----LeuSerThrHisTyrAlaGluGlnGlyAsn 240
Db 9390 CAAGTAAGTCGTAGGTTCAATCAGTATAGTACACCAAAATTTCTGCTGGAACAAC 9449
QY 241 HisArgLeuPheGluAla-----GlnArgAsnLeu----- 250
Db 9450 GACAAACATTATGAGGCCCACTAGGAATAAATGGAGAGCAAAATCTGGATTCTTAAAC 9509
QY 251 -----PheAlaGlyValSerPheProGluPhe 259
Db 9510 ATTCTTTTAACAATTCCTGAAATGCGTCTACCTTACACAAATAATCACAATCTCTCCACTG 9569
QY 260 LysAspSerProArgTrpArgGlnThrGly----- 269
Db 9570 AAAGATTCTCTCTATGCGGAAAAACAGCGCTTGAAGGAATCTTGAAAAACGACAAGCAA 9629
QY 269 ----- 269
Db 9630 TCATTGTTTAAGTGTAAGGCTCAAGTATAAGAAAAAACAACACAGGCAATCCATCACA 9689
QY 270 -----IleSerValLeuAsnThrGluIleLysGlnValTyrAlaAspGlyMetGln 287
Db 9690 AATCTCTTGTGCTGCTGCTTTGTGAGTTTATCATCAGTCAGAGCATCAAAATCTTTGACAGGCAT 9749
QY 288 PheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeuLysAlaIleTyrGly 307
Db 9750 TTTGAAAAAAC-----AGAAACATGCAATTAGATTTTGTCCACCAATCTATAAT 9800
QY 308 SerAlaLys-----ArgValAsnLeuGluLys-----GluPheProGln 320
Db 9801 GAAACAAAAATTAAGTTTGATAAGTACAAAGCTGAAAAATCTCACGACGAGCTCCCCAGG 9860
```

```
QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
DB 9861 ACCTTT-----
QY 341 TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGln 360
DB 9879 TTTTCTTCCAGTTCCAGTCTC-----AATGTTGAAGTGTCTCCA 9914
QY 361 Phe-----AlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleIleTyr 376
DB 9915 TTTACCATAGAGATGTCGGCAITTCGGCTATGTGTTCCCAAAACAGCTCAGCATGCTAGT 9974
QY 377 PheAlaThrAspGlyLysGlnGlyLysAlaProAsnPhe----- 389
DB 9975 TTTCTCCATCTAGTTCTCAGCTGCGTGTGCTTCATACACATTATTCGTCATCATTA 10034
QY 389 ----- 389
DB 10035 GAGCTGCCAGTCTCTCATGTCCTAGAAATCTCAAGCTTTCTCTTCCACATTTCAAGGA 10094
QY 390 -----LeuSerLysAlaLeuSerAsnAlaGlyPhe---TyrThr 401
DB 10095 TTGTGTACCATAGCCATATTTTATCTGTCGATGGCAATATATACCTATGATTTCTCC 10154
QY 402 PheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProProGly 421
DB 10155 TTTAAATCAAGT-----GTTCATCACATGTAATACCAATGCT 10190
QY 422 GluPheHisAlaGlnProaspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPhe 441
DB 10191 GAACTTTTAAACCATCATGAT-----ATTGTTGTCTCATCTCTCTTCTTCATCTTCATCT 10244
QY 442 ThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIle----- 456
DB 10245 GTTCATGTGACTCGACTGACATAAATAGAGGCGCACCAAGATTCACAAAGAAAAGGGA 10304
QY 457 -----MetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSer 470
DB 10305 TTGAAGTTAGCCACAGCTCTGTCTCTGAGCAAAATTTGTGAGGGTAGTCATAACAGT 10364
QY 471 ThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGlu 490
DB 10365 ACTGTGAGCTTAACACCGAAAATATGAAGTGTCAAGTGGCAAAACACCAACCAAGCCGAA 10424
QY 491 -----ThrGlyAsnAsnLeuAspValLeu 498
DB 10425 ATTCCAATTTTCAGATGAATTTCAAGCAAGNACTTAATGGAATATACCAAGTCAAAACCT 10484
QY 499 ThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsn 518
DB 10485 ACTGTCTCTCTCTCCATGGAATTTAAGTATGATTTCAATTTCTCAATGCTGTACTCTACC 10544
QY 519 LysLys-----TyrPheLeuVal 524
DB 10545 GCTAAAGAGAGAGTTGACACCAAGCTTAGCTTGGAAAGCCTCACCTCTTACTTTCCAT 10604
QY 525 IleAspArgAlaIleGlyGluAlaThrGlyAsnLeu----- 536
DB 10605 GAGTCATCTACCAAGGAGATCTCAAGGTTTCGGTTCTTCTCGGGAATATTCAGGAAC 10664
QY 536 ----- 536
DB 10665 ATTGCTAGTGGGCCAACACTTACTTGAATTTCCAGAGCACACGGTCTTCAGTGAAGCTG 10724
QY 537 ---GlyValHis-----TrpGlnLeuLysGluAspSerAsnProValPhe 550
DB 10725 CAGGCACTTCCAAATTCATGATATCTGGAACTTTGAAGTAAAGAAAATTTTCTCGGA 10784
QY 551 AspLysThrLysAsnArgValTyrThrTyrArgAspGly-----AsnAsnLeuMet 568
DB 10785 GAAGCCACACTCCAAACGCATATATTCTCTCTGGGAGCACAGTACGAAAACCACTTACAG 10844
```

```
QY 569 IleGlnSerLeu-----AsnAlaAspArgThrSer-----LeuAsnGluGlu 582
DB 10845 CTAGAGGGCTCTTTTTCACCAACGGAGAACTATACAGCAACCCCTGGAACTCTCT 10904
QY 583 GluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheValPheGlu 602
DB 10905 CCATGGCAAAATGTACGCTCTTGTTCAGGTCCATGCAAGTCAGCCAGTTCTTCCATGAT 10964
QY 603 LysPro-----LysLysAsnAlaGlyThrGlnAsn----- 612
DB 10965 TTCCCTGACCTTGGCCAGGAAGTGGCCCTGTAATGCTAACTAAGAACAGAGATCAGA 11024
QY 613 PheValSerIleValTyrProTyrAspGlyGlnLysAlaProGluIleSerIleArgGlu 632
DB 11025 TGGAAATGAAGTCCGATTCATTTCTGGGTCTTTCAGAGCCAGTCCAGCTT----- 11078
QY 633 AsnLysGlyAsnAspPheGluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGly 651
DB 11079 -----TCCATGATCAACCAAGGACACCTTGACATTCAGAGATCCTTAGAAGGA 11129
```

RESULT 5

US-09-801-368-369

; Sequence 369, Application US/09801368

; Patent No. US20020128250A1

; GENERAL INFORMATION:

; APPLICANT: Busby, Robert

; APPLICANT: Cali, Brian

; APPLICANT: Hecht, Peter

; APPLICANT: Holtzman, Doug

; APPLICANT: Madden, Kevin

; APPLICANT: Maxon, Mary

; APPLICANT: Milne, Todd

; APPLICANT: No. US20020128250A1man, Thea

; APPLICANT: Royer, John

; APPLICANT: Salama, Sofie

; APPLICANT: Sherman, Amir

; APPLICANT: Silva, Jeff

; APPLICANT: Summers, Eric

; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi

; FILE REFERENCE: 109272.147

; CURRENT APPLICATION NUMBER: US/09/801,368

; CURRENT FILING DATE: 2001-03-07

; PRIOR APPLICATION NUMBER: US 09/487,558

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: US 60/160,587

; PRIOR FILING DATE: 1999-10-20

; NUMBER OF SEQ ID NOS: 440

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 369

; LENGTH: 3996

; TYPE: DNA

; ORGANISM: Saccharomyces cerevisiae

US-09-801-368-369

Alignment Scores:

Pred. No.:	0.0378	Length:	3996
Score:	115.50	Matches:	157
Percent Similarity:	31.21%	Conservative:	108
Best Local Similarity:	18.49%	Mismatches:	287
Query Match:	3.31%	Indels:	297
DB:	11	Gaps:	45

US-09-802-285A-2 (1-659) x US-09-801-368-369 (1-3996)

```
QY 1 MetThrThrLysIlePheLysArgIleIleValPheAlaValIleAlaLeuSerSerGly 20
DB 73 ATCTCAAGTAAGTCTCGTGAAGAGATTATCAAGTCATTCAGCCATAAATCAAGATCT 132
QY 21 AsnIleLeuAlaGlnSer-----SerSerIleThrArg 31
DB 133 GATTGAAAGCATTTGGGTGGTTCGGAACAATAAGACGCGCCAGTCAGTGTACTTTT 192
QY 32 LysAspPheAspHisIleAsnLeuGluTyrSerGlyLeuGluLysValAsnLysAlaVal 51
```

```

193 193  AAGGAC---CGATACGTTTCAATGAATCGCTATACCTTGAAGAAAGCTAAAAAGACCGCT 249
      |||||  : : : : :  |||  |||: : : |||: : : |||
52  52  AlaAlaGlyAsnTyrAspAspAlaAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSer 71
      |||  : : : : :  |||  |||: : : |||: : : |||
250 250  TTAGATGACTACTACACGAGGGGCATAAACTCACTAACCGCTACGAGGAAGACGCGT 309
      |||  : : : : :  |||  |||: : : |||: : : |||
72  72  LysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnProIle 91
      |||||  : : : : :  |||||  |||||: : : |||||
310 310  GATGACGAAATATTTCGGTTGCTCTAATGGC----- 339
82  82  AspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLys 111
      |||  : : : : :  |||  |||: : : |||: : : |||
340 340  GACAGAAATGATGACGACCTG-----CACTCA 366
112 112  GlyTyrGlyTyrPheAsp-----TyrGlyLysAspIleAsnTrpGlnMetTrpPro 128
      |||  : : : : :  |||  |||: : : |||: : : |||
367 367  GGTGTCGAAGTTTCTCCACTACACCTTATTGCGAGAAATGAGG----- 411
129 129  ValLysAspAsnGluValArgTrpGln-----LeuHisArgValLysTrpTrpGln 145
      |||  : : : : :  |||  |||: : : |||: : : |||
412 412  TCAGACAGTGAATGACTAGCTTGAATGAAATGCGACCGAATGCTCAATGGCAGTCA 471
146 146  AlaMetAlaLeuValTyrHis-----AlaThrGlyAspGluLysTyrAlaArgGlu 162
      |||  : : : : :  |||  |||: : : |||: : : |||
472 472  ATGCTGGCCAGAGTGTGAAGGAGAGATATTGTTAAAGGTGAAGACGAGGATGCTAAC 531
163 163  TrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsn 182
      |||  : : : : :  |||  |||: : : |||: : : |||
532 532  CRAAGTC-----AAGAAACCA---GGGTAAATAAGGAGCTC 564
183 183  AspLysPheValTrpArgProLeuGlu-----ValSerAspArg 195
      |||  : : : : :  |||  |||: : : |||: : : |||
565 565  TCAGATGAGTATGGCTCGAATGAGGCGATGCTGAATGGGAGGACCATGCAAGAGATG 624
196 196  ValGlnSerLeu-----ProProThrPheSerLeuPheValAsnSerProAla 211
      |||||  : : : : :  |||||  |||||: : : |||||
625 625  GAACAGTCGCTTACATATTTAAGAGATAGTTCAGATTCGGTTTGAAGAGATAATGAAG 684
212 212  Phe-----ThrProAlaPheLeuMetGlu 219
      |||||  : : : : :  |||||  |||||: : : |||||
685 685  TTTCAAATTCACAGGCGCAAGATATTGAGCTGGATGCACTGGAGGCCATCTTACAAAG 744
220 220  PheLeuAsnSerTyrHisGlnAlaAspTyr-----LeuSerThrHisTyrAla 236
      |||||  : : : : :  |||||  |||||: : : |||||
745 745  CTATGAACAGATATCACCGGTGCTCTTATTGGCTTAACITGAATAAAATGTATAAG 804
237 237  GluGln-----GlyAsnHisArgLeuPheGluAlaGln----- 247
      |||||  : : : : :  |||||  |||||: : : |||||
805 805  GATAAACCAATACCAATACTGCAGAAATTTACCGCTAGATAAGACGTAATGAATCTTGG 864
248 248  -----ArgAsnLeu 250
865 865  CTGAACCTTTAAACGAACTTAACGTTTGAGGAGCAAGAGTTGAGCAGCTGGATAAAACCGT 924
251 251  PheAlaGlyValSerPhe-----ProGluPheLysAspSerProArgTrp 265
      |||||  : : : : :  |||||  |||||: : : |||||
925 925  TTCTCACCATAAGTAGTTCGGATAATTGCCAAGAGATTTTGAATGGTGGCCCAATGG 984
266 266  ArgGlnThrGlyLysSerValLeuAsnThrGluIleLysLysGlnValTyrAlaAspGly 285
      |||  : : : : :  |||  |||: : : |||: : : |||
985 985  ---AACTGCAAAATGAAGATTCCTGCAGAACAAATTGATGAAGGAAGAAACATCGAGTCT 1041
286 286  MetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeuLysAla 305
      |||  : : : : :  |||  |||: : : |||: : : |||
1042 1042  ATA---TTCCAAAAAATAATTTCTATCCGCTATCATCCTTGGATGTTCAAACTGAAACTA 1098
306 306  TyrGlySerAlaLysArgValAsnLeuGluLys----- 316
      |||  : : : : :  |||  |||: : : |||: : : |||
1099 1099  CATTTTATAGTCTACAGAGAACTTTGACAAAGATGAACATAAAATATCCTTTATGAAGG 1158
317 317  -----GluPhePro-----Gln 320
      |||||  : : : : :  |||||  |||||: : : |||||

```

```

1159 1159  TTAAGATCACTACTGGGTTCCCGTCTATTATTAATCAAGAAAGTATTTTGCATGATTG 1218
321 321  SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
      |||||  : : : : :  |||||  |||||: : : |||||
1219 1219  TCATATGCACGAAAGCTTAAATAATCAACAATGATGATGATCGATCAAAATGATCGATGAT 1278
341 341  TyrAsnThr----- 343
1279 1279  TTTAACGCTTTTANTTCGATCTTCTGTGCAATTGAAGTACACACTGACAAATAATGTCTCC 1338
344 344  -----ProMetPheGlyAspSerTrpIleThrAspLys 354
1339 1339  AATTTCGCGTTCGATGCGATTGACCCGACGCTTCGAAATACTGTAATA----- 1389
355 355  AsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaIleGlnAlaIle 374
      |||||  : : : : :  |||||  |||||: : : |||||
1390 1390  -----GAAGCCATT 1398
375 375  LysTyr-----PheAlaThrAspGlyLysGlnGly---Lys 385
1399 1399  CGTTATTTATTTTCTGTGTAATTTAAAGTTGATTCAGTCCAGTAAACAAATTTCAAA 1458
386 386  AlaProAsnPheLeuSerLys-----AlaLeuSerAsnAlaGlyPheTyrThrPhe 402
      |||||  : : : : :  |||||  |||||: : : |||||
1459 1459  GCACCCGATCTACTCTTGAATACTGGGATCACCTAAACAAACACCGGTCACTATATT--- 1515
403 403  ArgSerGlyTrpAspLysAlaThrValMetValLeuLysAlaSerProProGlyGlu 422
      |||||  : : : : :  |||||  |||||: : : |||||
1516 1516  -----AACGGTCAGAAACCGTGATT-----CCAAATGAA 1545
423 423  PheHisAlaGlnProAspAsnGlyThrPheGluLeu-----PheIleLysGlyArg 439
      |||||  : : : : :  |||||  |||||: : : |||||
1546 1546  TTCTCAAGTTTAACCTTTGAGACTCGTACATAATAATTCATCTTTTGAACAAACAA 1605
440 440  AsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeu 459
      |||||  : : : : :  |||||  |||||: : : |||||
1606 1606  AACTTCCCACCAACATTGCT-----AACGCTCAGAAGCA----- 1641
460 460  ArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMet 479
      |||||  : : : : :  |||||  |||||: : : |||||
1642 1642  GAAAAATGG-----CTAAGTTCCATTTCGAAAAATTTG 1674
480 480  ValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThr 499
      |||||  : : : : :  |||||  |||||: : : |||||
1675 1675  GGTGCCATGAAAGAAAGCTGAACAGTTG-----AGCAATATTCTGACAGGCG 1725
500 500  TyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLys 519
      |||||  : : : : :  |||||  |||||: : : |||||
1726 1726  TTCCAAAATTCGCTGTTTATCAGATTATCATTAATGACACACTTGT-----AAA 1776
520 520  Lys-----TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGly 534
      |||||  : : : : :  |||||  |||||: : : |||||
1777 1777  AAGTAAAAAGATGCTCACTATTCTTGGTATATCTCC-----GGTAACACTTTTGAG 1827
535 535  AsnLeuGlyValHisTrp-----GlnLeuLysGluAspSerAsnProVal 549
      |||||  : : : : :  |||||  |||||: : : |||||
1828 1828  TCTAGTGGTGATATATGTTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1887
550 550  PheAspLysThrLysAsnArg-----ValTyrThrThrTyrArgAspGly 564
      |||||  : : : : :  |||||  |||||: : : |||||
1888 1888  TTAAGAATTTTCGAAATAAATCCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1947
565 565  AsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn----- 580
      |||||  : : : : :  |||||  |||||: : : |||||
1948 1948  AATAATTTCAATGTGTATGATATACACAAAGAAACAGATTTGAACATCTCTAGTATCG 2007
581 581  -----GluGluGluGlyLysValSerTyr-----ValTyrAsnLysGluLeuLys 595
      |||||  : : : : :  |||||  |||||: : : |||||
2008 2008  AAGGGGAGGATTCGAAAGAAATTCCTTACTACCGAGTAGTAGCAAAATTCGTCAAGTGAT 2067
596 596  ArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGlnAsnPhe----- 613
      |||||  : : : : :  |||||  |||||: : : |||||
2068 2068  TTGGACAGCATGCTCATCAGTCCAAAGAAAGAAATTTTTCACAGACCTTTTGATCAG 2127

```

```

QY      15  IleAlaLeuSerSerGlyAsnIleLeuAlaGln---SerSerSerIleThrArgLysAsp  33
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     1360  ATAGCAGAGATACAGGGAAAATGTTGGCTCAAAATGGAGACGGCTCTTACTCGAACGAAG  1419

QY      34  PreAspHisIleAsnLeuGluTyrSerGlyLeuGluLysValAsnLysAlaValAlaAa  53
      ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     1420  GACGAGCAAAATT-----AAAATTTCAAGGAAGCTGGAAGGAA  1455

QY      54  GlyAsnTyrAspAspAlaAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAla  73
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     1456  GAATCCTCTAACTTATCTGCTAAACAGCGTGCTGCTGTGTGAAAAATAGTATATAAGCA  1515

QY      74  ArgGluProAspPheSerAsnAlaGluLysProAlaAsp-----  86
      ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     1516  AAAAGAAAAACAATAAAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAT  1575

QY      87  -----IleArgGlnProIleAspLys-----Val  94
      ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     1576  GATCAGTATTATGTAAAGGACCAAAATTCAAGAAAGACGACGATCACAGTTGGAAAAA  1635

QY      95  ThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGly-----  112
      ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY 414 ValLeuLysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGlu 433
DB 2572 -----AAAGCAGGCCATCTTACAGGGTTGATCGACAGCATCCGACGACATTT--- 2622
QY 434 LeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAsp 453
DB 2622 ----- 2622
QY 454 GluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThr 473
DB 2622 ----- 2622
QY 474 LeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsn 493
DB 2623 -----AACGCCAATAAGTTCTCGACATAAT 2649
QY 494 AsnLeuAspValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSer 513
DB 2650 AACATT-----TTGAATCG-----ATAGATAATAGTTGGCT 2682
QY 514 ValLeu-----PheIleAsnLysTyrPheLeuValIleAspArgAla----- 528
DB 2683 GCAATTAACTATATCAAGCGCATATGAGACATTAGCAATCCAGGCTTAAATCA 2742
QY 529 -----IleGlyGluAlaThrGlyAsnLeu----- 536
DB 2743 ATGCCCGAGGTGGCGCTATGTGGTTATGCAAAAGGTGGTATCTCACCCTGGACGGGT 2802
QY 537 GlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArg 556
DB 2803 GGCTCAAAATGGCCATCTTAATACAGAGG-----TATGATGAACAACAGATCAGC 2856
QY 557 ValTyrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArg 576
DB 2857 GAAGATCCCTCATACAGGAGCGTAATATTGGGTATGGCTCGGATGGTAGCAAACTT 2916
QY 577 ThrSerLeuAsnGlu---GluGluGlyLysValSerTyrValTyrAsnLysGluLeuLys 595
DB 2917 GCGTACTTCGGATTACAGGATGGGATGATTCA-----AAAGCACTACTC 2964
QY 596 ArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSer 615
DB 2965 CTGCTCAAAAGTTTCGAAAGCCCAACAGCAGCTACCACTAATGATTTTCTGTA 3024
QY 616 IleValTyrPro-----TyrAspGlyGlnLysAla 625
DB 3025 GACATCAG-CCGTGTTCTGAAGAATCAAGAAACAAATCAGTATGATGGC----- 3074
QY 626 ProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLysGlyLys---LeuAsn 644
DB 3075 -----AAAGCAATTTGATTTTCTAAATAAAATGTCAGCT 3110
QY 645 LeuThrLeuThrIleAsnGlyLysGlnGln 654
DB 3111 CTTGCACGACTTTTACTAAAGACACAA 3140

RESULT 7

US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: P8186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; PRIOR FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787

; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3921)..(3921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature

```
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
```

```
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)
```

Alignment Scores:

Pred. No.:	590	Length:	1830121
Score:	115.00	Matches:	139
Percent Similarity:	32.80%	Conservative:	85
Best Local Similarity:	20.35%	Mismatches:	203
Query Match:	3.29%	Indels:	256
DB:	15	Gaps:	41

US-09-802-285A-2 (1-659) x US-10-329-960-1 (1-1830121)

```
QY 34 PheAspHisIleAsnLeuGluTyrSerGlyLeuGluLysValAsnLysAlaValAlaLa 53
Db 676209 TGGGATCATATTAAACTAAGTTATTCTTCACAAAAAATCACCATAAAGCA----- 676159
QY 54 GlyAsnTyrAspAlaAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAla 73
Db 676158 -----CGTTCTCATGAATATTGT 676141
QY 74 ArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnProfileAspLys 93
Db 676140 CATCAATCCACTTGTAATGGTGAAGTAACCCCAAGGATTGCATTAGTAGAAGAAAAA 676081
QY 94 ValThrArgGluMetAlaAspLys-----AlaLeu 103
Db 676080 GGGGTTTATAAAATTAAGGATAAATATGAGGTGAACTTGAAAGCAAAAGAAATTGGCTGG 676021
QY 104 ValHisGlnPheGlnProHisGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsn 123
Db 676020 TCTCATGAATTTAAGAAATAGTAGGGTCAA-----GATGCTGATAAGGATATTCTCA 675970
QY 124 TrpGlnMet-----TrpProValLysAspAsnGluValArg 135
:::
```


NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature

LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)


```

/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (145058)..(145058)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (145171)..(145171)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (145942)..(145942)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (147197)..(147197)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (150841)..(150841)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (152500)..(152500)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (152530)..(152530)

```

QY	134	valArgTrpGlnLeuHisArgValLysTyrTrpGlnAlaMetAlaLeuValTyrHisAla	153
Db	2230	GTGAGTGGACCATTAAGTACGTA-----	2253
QY	154	ThrGlyAspGluLysTyrAlaAaGluTrpValTyrGlnTyrSerAspTrpAlaAaGlyLys	173
Db	2254	---GGGACACGGAGTACAGAACTTTTGAGATCTATGCAACTGAGCCATGGATTAGCCCT	2310
QY	174	AsnProLeuGly-----LeuSerGlnAspAsnAspLysPheValTrpArgProLeuGlu	191
Db	2311	TTTGTTCAGTGGAAAGTGTAATCTTAGAGACAATACCGAGTTGTC-----CTTAGG	2361
QY	192	ValSerAspArgValGlnSerLeuProProThrPhe-----SerLeuPheValAsn	208
Db	2362	GTGAATATGATGTAGAGGGTCTTGACCCAGGTCTCTATGTTGAAGGATAATCATTTGAT	2421
QY	209	SerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSer-----	223
Db	2422	GATCCCAACA--ACGCCAGTTATTGAAGACGAGATCTTGAACACAATTGTTATTCOCGAG	2478
QY	224	---TyrHisGlnGlnAlaAspTyrLeuSerThrHisTyr-----AlaGluGln	238
Db	2479	AAGTTCACCTCTCAGAACCAATTAACCCCTCACCTGGTATGATTAATGTCGCAGAAATG	2538
QY	239	GlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGlu	258
Db	2539	GTGACTCACCACCTTTCATCTGTCCTGAGGGAGTGAGCTTCTTACGGGATGCCACA	2598
QY	259	PheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeu-AsnThrGlu----	276
Db	2599	TACTGGGACTACCGTCTGTACAGACACAGATGGAATGTTGTGTCCATACCCAGCTAGAT	2658
QY	277	-----IleIysLysGlnValTyrAla-----AspGlyMetGlu	287
Db	2659	TATCTTCCCGTGGAGTCTCAATTCOAATCCCTGGAAACTGGGAGCTAGTATGCAGCTGGA	2718
QY	287	npHeGluLeuSerProIleTyrHisValAlaAlaIleAspIlePhe-----	302

```

QY 591 -----AsnLysGluLeuLysArgProAlaPheValPheGluLysProLysLysAs 607
Db 3687 TGAATACTAGAGAGCAATAATAGCCCAAGCTTGCATATCATCACCAAGCAAA 3746
QY 607 nAlaGlyThrGlnAsnPheValSerIleValTyrProTyrAspGly-----GlnLysAl 625
Db 3747 TGCT-----ACCATAGTATCAGTTGAGATCGAGAGCTGAGGTGGCCTTAAAGAC 3797
QY 625 aProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLysGlyLysLeuAsnLe 645
Db 3798 AGTGACAGTGGAAATACTATAAC-----GGAAACCGCTAATGA 3836
QY 645 uThrLeuThrIleAsnGlyLysGlnGlnLeuValLeuValPro 659
Db 3837 GATGCAACAATA-----GTGGTTCCTGTTCCCT 3864

RESULT 10
US-09-070-927A-321
; Sequence 321, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 321:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6720 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 321:
US-09-070-927A-321

Alignment Scores:
Pred. No.: 0.185 Length: 6720
Score: 112.50 Matches: 109
Percent Similarity: 37.41% Conservative: 93
Best Local Similarity: 20.19% Mismatches: 202
Query Match: 3.22% Indels: 138
DB: 11 Gaps: 28

```

```

US-09-802-285A-2 (1-659) x US-09-070-927A-321 (1-6720)
QY 145 GlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpVal 164
Db 2195 CAGGCATACGCGTTAATA-----AAAGAAAAATGCTATTTCACAAAAATGCTGC 2242
QY 165 TyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLys 184
Db 2243 AGTCTGCTTCTTCTGATTATTTCAGAGGAA-----GGCGAACCAAAATCTACTAATAGG 2296
QY 185 PheValTyrArgProLeuGluValSerAspArgValGlnSerLeuProProThrPheSer 204
Db 2297 AATGTTTGGCGACACCTTGATGTGGATTCGAATTCGAGTGAATCGGATCGAAGAGTTTGACT 2356
QY 205 LeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeu---AsnSer 223
Db 2357 ---TATATTCTATTGCTGATTATAAACAATGGGAATTGATAAAGTGTTCGGAACGCT 2413
QY 224 TyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeu 243
Db 2414 TTGCTGTTAATCTCGAGTATTTTGGAGAGGAGCTATATCGATAAG-----TACAGACTT 2467
QY 244 PheGlu-----AlaGlnArgAsnLeuPheAlaGlyVal 254
Db 2468 AGTAATTTGGGAGTATTGTTAAACGGGTGGAGTGGCTGCTATGATCTATT----- 2518
QY 255 SerPheProGluPheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsn 274
Db 2519 ---CTTCCAGAACTAGTGAACAGAGATAAATAATATATGGGACCGGTTA-AC TGACTTA-- 2571
QY 275 ThrGluIleLysLysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyr 294
Db 2572 -----CAATTTTATTCAGATGGCATTTCATTTGGGAACAGACGCGATTGTAT 2616
QY 295 HisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu 314
Db 2616 ----- 2616
QY 315 GluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleAlaLeuIle 334
Db 2617 CAATATGAAGTTTGTGATGACGTATGATATCTATTACAG-----ATTCTGAATATCTT 2670
QY 335 SerIleSerLeuPro---AspTyrAsnThr-----PromethGlyAspSer 349
Db 2671 GAAATATCATTCGCTTGGATCTTCGACACGAAATGAAGATACCTATCTATCTACCTAC 2730
QY 350 TrpIleThrAspLys-----AsnPhe 356
Db 2731 TATATAACGGATAAATCAAGATGTTCTGAATCCTATCAACGATAGCGATCATGTCAATTT 2790
QY 357 ArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyr 376
Db 2791 CGATAT-----GTATATGATAGCTATCTGTAATATGAAG-- 2823
QY 377 PheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsn 396
Db 2824 ---AAAGAACTTGGAAAAACACTGCAGCCAGAGAAACTTTTTCGTGGTGAATCAAGTGGT 2880
QY 397 AlaGlyPheTyrThrPheArgSerGlyTyrAspLysAsnAlaThrValMetValLeuLys 416
Db 2881 TTAGTGTGTTTATAAACACAGAGGAT-----ATTTATTTTAGCTTTT 2922
QY 417 AlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIle 436
Db 2923 AATGGATTGCAATGGAAGCTCACATGGTCTATGCTTCTACAGGGAGTTTACGCTACAATTA 2982
QY 437 LysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIle 456
Db 2983 CAGGGTGATGATTTGATATCTGATAGTGGTGTGTACAGTTAT----- 3024
QY 457 MetLysLeuArgAsnTrpTyr-----ArgGlnThrArgIleHisSerThrLeuThrLeu 474
Db 457 MetLysLeuArgAsnTrpTyr-----ArgGlnThrArgIleHisSerThrLeuThrLeu 474

```

Dbb 3025 GTCAACAAAGCAGAGTGGCTTCAACCGAAGGAGTGTGATTGCGCATATAACAATGTTTATT 3084
QY 475 ---AspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGly--- 492
Dbb 3085 GCAGAGAATTCATCACTTACTAGT-ACT-----GATACATGGGGCTACGGGAAA 3131
QY 493 -----AsnAsnLeuAspValLeuThrTyrThrAsnPro 503
Dbb 3132 TTACCAACCCCATTTTTCAGCGAATAAAGAAATTCCTGTCGGTTTTTTTGGGAATGC 3191
QY 504 SerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLysTyrPhe--- 522
Dbb 3192 GGCTGGATAGATAAAGATGATCAGAACTTAAATGATTTT-----AAGCGTAGTTTATC 3245
QY 523 -----LeuValIleAspArgAlaIleGlyGluAlaThrGly 534
Dbb 3246 TACTTCAAGCCATTAATTCAGTAATATTATGATAGTTTATAGTCAAAAAGAGACT 3305
QY 535 AsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAspLysThrLys 554
Dbb 3306 GAAATTACGAGTACTTATAATTAGCACCTACGCTAAAT-----TGTCAAAAAAGCAGA 3359
QY 555 AsnArgValTyr-ThrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAl 574
Dbb 3360 TCAATTCAATTTAACTACAAATTCACAAAATATAAACTCCTTATAGTATCAGGTCAAAAC 3419
QY 574 aAspArgThrSerLeuAsnGluGluGluGlyLysValSerTyrValTyrAsnLysGluLe 594
Dbb 3420 GGATCAAACTACTGTGAAAGATCAGAG-----ATTATAATCAGCTGAT 3464
QY 594 ulysArgProAlaPheValPheGluLys-----ProLysLysAsnAl 608
Dbb 3465 CGAA-----CTATATACGCTTCACCAATAAATTTTCAATATAAAGCGGGAAGAGTTC 3518
QY 608 aglyThrGlnAsnPheValSerIleValTyrPro---TyrAspGlyGlnLysAlaPro 626
Dbb 3519 AGGATACCGTGATTTCCCTTTGGAGAGATATCCAGATTATGCCAGTAAAGTGAACCA 3576

RESULT 11

US-09-993-292A-19
; Sequence 19, Application US/09993292A
; Patent No. US20020146430A1
; GENERAL INFORMATION:
; APPLICANT: University of Maryland
; TITLE OF INVENTION: USE OF CLY A HEMOLYSIN FOR EXCRETION OF
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: UOFMD.007A
; CURRENT APPLICATION NUMBER: US/09/993,292A
; CURRENT FILING DATE: 2001-11-23
; PRIOR APPLICATION NUMBER: 60/252,516
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 2253
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ClyA::SacB fusion gene
; NAME/KEY: CDS
; LOCATION: (0)...(2253)
US-09-993-292A-19

Alignment Scores:
Pred. No.: 0.0501 Length: 2253
Score: 111.00 Matches: 124
Percent Similarity: 32.18% Conservative: 80
Best Local Similarity: 19.56% Mismatches: 196
Query Match: 3.18% Indels: 234
DB: 11 Gaps: 34

US-09-802-285A-2 (1-659) x US-09-993-292A-19 (1-2253)

QY 55 AsnTyrAspAsp-----AlaAlaLysAlaLeuLeuAlaTyr 66
Dbb 793 GATTATGATGATTAAATGCTTTCTTTTAAAGAGAGCTGCAAGAAAATGATTAAACACC 852
QY 67 TyrArgGlu-----LysSerLysAlaArgGluProAspPheSer 79
Dbb 853 TGTAAATGATACCAACAACAGCTCATGGTAAGAAGACGCTTTTCGAGGTTCTCGACGTCGT 912
QY 80 AsnAlaGlu-----LysProAlaAspIleArgGlnProIleAspLysValThrArg 96
Dbb 913 AGTAAAGAAACCAAAAGCCATATAGGAAACATACGCGCATTTCCCATATTATACACGC 972
QY 97 GluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPhe 116
Dbb 973 CAT-----GATATGCTGCAAATCCTGAAACAGCAAAAATAATCAAAATATCAAGTTCCT 1026
QY 117 AspTyr-----GlyLysAspIleAsnTrp 124
Dbb 1027 GAATTCGATTGCTCCACAATTAATAATATCTCTTCTGCAAAAGGCGCTGGACGTT--TGG 1083
QY 125 GlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrp 144
Dbb 1084 GACAGCTGGCCATTACAAAACGCTGAC-----GGCACTCTCGCAAACTAT 1128
QY 145 GlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrp--- 163
Dbb 1129 CAGGCTACCATCGTCTTTGTCANTAGCCGAGATCCTAAAAATGCGGATGACACATCG 1188
QY 164 ValTyrGlnTyr-----SerAspTrpAlaArgLysAsn 174
Dbb 1189 ATTACATGTTCTATCAAAAGTCGCGCAAACTTCTATTGACAGCTGG-----AAAAAC 1242
QY 175 ProLeuGlyLeuSerGlnAspAsnLysPheValTrpArgProLeuGluValSerAsp 194
Dbb 1243 GCTGCGCGCTCTTTAAAGACAGCAAAATTCGATGCAAAATGATTCTATCTCTAAAGAAC 1302
QY 195 ArgValGlnSerLeuProThrPheSerLeuPheValAsnSerProAlaPheThrPro 214
Dbb 1303 CAACACACAGAA-----TGGTCAGGTTTACGCCACATTTACATCT 1341
QY 215 AlaPheLeuMetGluPheLeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHis 234
Dbb 1342 GACGGAAAAATCCGTTTATTC-----TACACTGATTCTCCGCTAAACAT 1386
QY 235 TyrAlaGluGlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyVal 254
Dbb 1387 TAC-----GGCAACAAACACTGCAACTGCACAAAGTTAACGATCATCAGCATCA--- 1434
QY 255 SerPheProGluPheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsn 274
Dbb 1435 -----GACAGCTCTTGAACATCAACGCT 1458
QY 275 ThrGluIleLysLysGlnValTyr---AlaAspGlyMetGlnPheGluLeuSerProIle 293
Dbb 1459 GTAGAGGATTATAAATCAATCTTTGACGGTGACGGA----- 1494
QY 294 TyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsn 313
Dbb 1495 -----AAAAAGTATCAAAATGTACAGCAGTTCATC 1524
QY 314 LeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeu 333
Dbb 1525 GATGAAGGCAACTACAGCTCAGCGCACCAACCATACGCTGAGAGAT----- 1569
QY 334 IleSerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAsp 353
Dbb 1570 -----CCTCACTAC-----GTAGAAGAT 1587
QY 354 LysAsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAla 373
Dbb 1588 AAAGGCCACAAATACTTA-----GTATTTGAAGCAACACTGGA 1626

```
QY 374 IleLysTyrPheAlaThrAspGlyLysGlnGlyLysAlaProAsnPhelLeuSerLysAla 393
Db 1627 ACTGAA- - - - -GATGGTACCAGCGAA- - - - -GAATCT 1656
QY 394 LeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTyrAspLysAsnAlaThrValMet 413
Db 1657 TTATTTAACAAGCATACTATGCGAAAGCAGCATCATCTTCGTCACAGAAAGTCAAAA 1716
QY 414 ValLeuLysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGlu 433
Db 1717 CTTCTGCAAAAGGATAAA- - - - -AAACGACGGCTGAGTTAGCAAAACGGCGCTCTCGGT 1770
QY 434 LeuPheLysGlyArgAsnPhelThr- - - - - 442
Db 1771 ATGATTGACTAAACGATGATTACACACTGAAAGTGAAGAAACCGCTGATGTCATCT 1830
QY 443 - - - - -ProAspAlaGlyValPheValTyrSerGlyAspGluAla 455
Db 1831 AACACAGTAACAGATCAAAATTGACCGCGGACGCTTTTAAATGNAACGCG- - - - - 1881
QY 456 IleMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHis- - - - -SerThrLeuThrLeu 474
Db 1882 - - - - -AAATGGTACCTGTCTCACTGACTCCGCGGATCAAAAATGACGATT 1926
QY 475 AspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsn 494
Db 1927 GACGGC- - - - -ATTACGCTAACGAT 1947
QY 495 LeuAspValLeuThrTyr- - - - -ThrAsnProSerTyrProAsnLeuAsp 509
Db 1948 ATTACATGCTGGTATGTTCTTAATCTTTAACTGGCCCATACACGCG- - - - - 1998
QY 510 HisGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAspArgAlaIle 529
Db 1999 - - - - -CTGAACAAA- - - - - 2007
QY 530 GlyGluAlaThrGlyAsnLeuGlyHisTrpGlnLeuLysGluAspSerAsnProVal 549
Db 2008 - - - - -ACTGGCCTTG- - - - -TTAAATGGATCTTGATCCT- - - - - 2040
QY 550 PheAspLysThrLysAsnArgValTyrThrTyr- - - - -Arg 562
Db 2041 - - - - -AACGATGTAACTTTACTTACTACACTTCCTGCTGACCTCAAGCG 2085
QY 563 AspGlyAsnLeuMetIleGlnSer- - - - -LeuAsnAlaAspArg 576
Db 2086 AAAGGAACAATGCTCGTATTACAGCTATATGACAAACAGAGGATTCTACGCAGACAAA 2145
QY 577 ThrSer- - - - -LeuAsnGluGluGlyLysValSerTyrVal 589
Db 2146 CAATCAACGTTTGGCGCAAGCTTCTGCTGAAATCAATCAAGGCGCAAGAAACATCTGTGTC 2205
QY 590 TyrAsnLysGluLeuLysArgProAlaPheValPheGluLys 603
Db 2206 AAAGACAGCATCTTGAACAGGACAAATTAACAGTTAAACAAA 2247
```

RESULT 12

```
US-09-993-292A-18
; Sequence 18, Application US/09993292A
; Patent No. US2002014630A1
; GENERAL INFORMATION:
; APPLICANT: James E. Galen
; APPLICANT: University of Maryland
; TITLE OF INVENTION: USE OF CLY A HEMOLYSIN FOR EXCRETION OF
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: UOFMD.007A
; CURRENT APPLICATION NUMBER: US/09/993, 292A
; PRIOR FILING DATE: 2001-11-23
; PRIOR APPLICATION NUMBER: 60/252,516
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
```

```
; LENGTH: 8908
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pSEC84sacB vector
US-09-993-292A-18

Alignment Scores:
Pred. No.: 0.424 Length: 8908
Score: 111.00 Matches: 124
Percent Similarity: 32.18% Conservative: 80
Best Local Similarity: 19.56% Mismatches: 196
Query Match: 3.18% Indels: 234
DB: 11 Gaps: 34

US-09-802-285A-2 (1-659) x US-09-993-292A-18 (1-8908)

QY 55 AsnTyrAspAsp- - - - -AlaAlaLysAlaLeuLeuAlaTyr 66
Db 1308 GATTATGATGATTATGCTTCTTTATTAAAGAGGCTGCAAGAAATGATTAAACACC 1367
QY 67 TyrArgGlu- - - - -LysSerLysAlaArgGluProAspPheSer 79
Db 1368 TGTAAATGAATACCAACACGTCATGTAAGAAGACGCTTTTCGAGGTTCTCGACGCGCT 1427
QY 80 AsnAlaGlu- - - - -LysProAlaAspIleArgGlnProIleAspLysValThrArg 96
Db 1428 AGTAAGAAGAAACCAACCAAGCCATATAGGAAACATACGCGCATTTCCCATATTACACGC 1487
QY 97 GluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPhe 116
Db 1488 CAT- - - - -GATATGCTGCAAAATCCCTGACAGCAAAAATGAAAAATATCAAGTTCCT 1541
QY 117 AspTyr- - - - -GlyLysAspIleAsnTrp 124
Db 1542 GAATTCGATTCTCCACAAATATAAATAATCTCTCTGCAAAAGCGCTGACGCT- - - - -TGG 1598
QY 125 GlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrp 144
Db 1599 GACAGCTGGCCATTACAAACCGCTGAC- - - - -GGCACTGTGCAAACTAT 1643
QY 145 GlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrp- - - 163
Db 1644 CAGGGCTACCATCTCTTTCGATAGCGGAGATCTCTAAATGCGGATGACACATCG 1703
QY 164 ValTyrGlnTyr- - - - -SerAspTrpAlaArgLysAsn 174
Db 1704 ATTACATGTTCTATCAAAAAGTCGGCGAAACTTCTATTGACAGCTGG- - - - -AAAAAC 1757
QY 175 ProLeuGlyLeuSerGlnAspAsnAspLysPheValTrpArgProLeuGluValSerAsp 194
Db 1758 GTGGCCGCGCTCTTTAAAGACAGCGCAAAATTCGATGCAAAATGATTCTCTCTCTAAAAGAC 1817
QY 195 ArgValGlnSerLeuProThrPheSerLeuPheValAsnSerProAlaPheThrPro 214
Db 1818 CAACACACAA- - - - -TGGTCAAGTTCAGCCACATTATCATCT 1856
QY 215 AlaPheLeuMetGluPheLeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHis 234
Db 1857 GACGGAAAAATCCGTTTATTTC- - - - -TACACTGATTCTCCGCTAAACAT 1901
QY 235 TyrAlaGluGlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyVal 254
Db 1902 TAC- - - - -GGCAACACAAACACTGACAACTGACAAAGTTAAACATATCAGCATCA- - - 1949
QY 255 SerPheProGluPheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsn 274
Db 1950 - - - - -GACAGCTCTTTGAACATCAACGGT 1973
QY 275 ThrGluIleLysLysGlnValTyr- - - - -AlaAspGlyMetGlnPheGluLeuSerProIle 293
Db 1974 GTAGAGGATTATAAATCAATCTTTGACGCGTGACGGA- - - - - 2009
```

```
QY 294 TyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsn 313
Db 2010 -----AAACGTATCAAAATGTACAGAGTTTCATC 2039
QY 314 LeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeu 333
Db 2040 GATGAAGGCAACTACAGCTCAGCGCAACACCATACGCTGAGAGAT----- 2084
QY 334 IleSerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAsp 353
Db 2085 -----CCTCACTAC-----GTAGAAGAT 2102
QY 354 LysAsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAla 373
Db 2103 AAAGGCCACAAATACTTA-----GTATTTGAAGCAAACTGGA 2141
QY 374 IleLysTyrPheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAla 393
Db 2142 ACTGAA-----GATGGCTACCAAGCGAA-----GAATCT 2171
QY 394 LeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMet 413
Db 2172 TTATTTAACAAGCATACTATGCGAAAGCACATCATCTCCGTCAGAAAGTCAAAA 2231
QY 414 ValLeuLysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGlu 433
Db 2232 CTTCTGCAAGCGATAAA-----AAACGACGGCTGAGTAGCAACGGCGCTCTCGT 2285
QY 434 LeuPheIleLysGlyArgAsnPheThr----- 442
Db 2286 ATGATTGACTAAACGATGATTACACATGATGAAAAAGTGATGAAACCGCTGATTCATCT 2345
QY 443 -----ProAspAlaGlyValPheValTyrSerGlyAspGluAla 455
Db 2346 AACACAGTAACAGATGAATTAACCGCGCAAGCTTTAAATGAACGGC----- 2396
QY 456 IleMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHis---SerThrLeuThrIleu 474
Db 2397 -----AAATGGTACCTGTTCACTGACTCCGCGGATCAAAAATGACGATT 2441
QY 475 AspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsn 494
Db 2442 GACGGC-----ATTACGCTCAACGAT 2462
QY 495 LeuAspValLeuThrTyr-----ThrAsnProSerTyrProAsnLeuAsp 509
Db 2463 ATTTACATGCTGGTTATGTTCTAATCTTTAACTGGCCCATACAGCG----- 2513
QY 510 HisGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAspArgAlaIle 529
Db 2514 -----CTGAACAA----- 2522
QY 530 GlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnProVal 549
Db 2523 -----ACTGGCCTTGTG-----TTAAAAATGATCTTGATCCT--- 2555
QY 550 PheAspLysThrLysAsnArgValTyrThrTyr-----Arg 562
Db 2556 -----AACGATGTAACCTTACTTACTACACTTCGCTGCTACCTCAAGCG 2600
QY 563 AspGlyAsnAsnLeuMetIleGlnSer-----LeuAsnAlaAspArg 576
Db 2601 AAAGAAACAAATGTCGTGATTACAGCTATATGACAAACAGAGGATTTCTACGCACACAA 2660
QY 577 ThrSer-----LeuAsnGluGluGluGlyLysValSerTyrVal 589
Db 2661 CAATCAACGTTTGCCCAAGCTTCTGCTGAAATCAAGGCAAGAAACATCTCTTGTCT 2720
QY 590 TyrAsnLysGluLeuLysArgProAlaPheValPheGluLys 603
Db 2721 AAAGACAGCATCTTGAACAAAGGACAAATTAACAGTAACAA 2762
```

RESULT 13

```
US-09-815-242-4610
; Sequence 4610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4610
; LENGTH: 1689
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4610

Alignment Scores:
Pred. No.: 0.0365 Length: 1689
Score: 110.50 Matches: 105
Percent Similarity: 34.25% Conservative: 94
Best Local Similarity: 18.07% Mismatches: 211
Query Match: 3.16% Indels: 171
DB: 10 Gaps: 26

US-09-802-285A-2 (1-659) x US-09-815-242-4610 (1-1689)
QY 91 IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis 110
Db 22 ATCGAACCATTTTAATGAATATTTTACAGAGAAATTTATATCAAGTACCCTCAA 81
QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
Db 82 AAT-----CATCAATGGTCTATACAA 102
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAla---Leu 149
Db 103 TTATCTGAGCTCGAAACAGCTTAACAGGTCATTCGCTTATTCGTCGATGGGCATCAT 162
QY 150 ValTyrHisAlaThrGly-----AspGluLysTyrAlaArgGluTrpValTyrGln 166
Db 163 ATGTATCATTCAGAGTATGCTTATTCATGTTAAAGTAAGAAATTAACAACATATAA 222
QY 167 TyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheVal 186
Db 223 GAAGCCATTGCGCGTATTTTCGAACATATGCTCAAGTGCAGATAATCAAAACGCGCATG 282
QY 187 TrpArgProLeuGluValSerAspArgValGlnSerLeuProProThrPheSerLeuPhe 206
Db 283 CAACAACTATG----- 294
```

```
QY 207 ValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGln 226
Db 295 -----GGCAGAAATTATCTGCACATCGATATAGCATTCATCGC 333
QY 227 GlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGluAla 246
Db 334 ACGCGCGGTATTGTGCAAGATACACAATAGATACCGAGAGATCGCTATATCGTTTCA 333
QY 247 GlnArgAsnLeuPheAlaGlyValSerPhe---ProGluPheLysAspSerProArgTyr 265
Db 394 GAACAATCTTTATACCTTAGGTCTATCCATTCATCCGACTCCTAAGAGTGCAGGTGGTT 453
QY 266 ArgGlnThrGlyIleSer----- 271
Db 454 TCGAAGCAGATTAGAGAAATATGACCCGAAATGTCATACATCATTCCAATTGCATTAT 513
QY 272 ---ValLeuAsnThrGluIleLysLysGlnValTyrAlaAspGlyMetGlnPheGluLeu 290
Db 514 TTAGCTGTGCATCAAGATGTTCTCTCAGCGCTATGTAGAGGTAAAGAAGATCAGTT 573
QY 291 SerProIle---TyrHisValAlaAlaIleAspIle-----PheLeu 303
Db 574 GAGAAAGTGTGTATCAATTAGCAGATATATAGATATACAGATACCTTAAAGATTTTAT 633
QY 304 LysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrVal 323
Db 634 TTATTACCAATACATCTTATCAAAATCAATGTGTGCAACGATCCACAG---TATTG 690
QY 324 GlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnThr 343
Db 691 CAATATAGTGAACAAGGTTTAAATAAAGACCTTGGCGTTCC----- 732
QY 344 ProMetPheGlyAspSer----- 349
Db 733 -----GGTGATTTCAGTTTACCGAGCTCTTCGGTTAGAACTGTATTTTCAAAAGCA 783
QY 350 -----TrrPileThrAspLysAsnPhe---ArgMetAla 359
Db 784 TTAACAATTTATTAAATTTACCGATACAGTTTAAATTTACTAATTTTATACGTACGAAT 843
QY 360 GlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIle----- 374
Db 844 GATCTTGAACAGATTGAACGCAATCGATGCTGCACAAGTTATCGCATCAGTCAAAAGAT 903
QY 375 -----LysTyrPheAlaThrAspGlyLysGlnGlyLysAlaPro 387
Db 904 GAGGTGAAACACCCCATTTTAAATTTGATGTTTGAAGAAGGATATCGTGCATTTGTACCG 963
QY 388 AsnPheLeuSerLysAla-----LeuSerAsnAlaGlyPheTyr 400
Db 964 AATCCATTAGGGCAACAGTTGAACCTGAATGATTTATTAACAATAGTGCATGATT 1023
QY 401 ThrPheArgSerGly-----TrpAspLysAsnAlaThrValMetValLeu 415
Db 1024 GTT---CGTAGAGGATACCGAATTTACCATGCTGATAAAGATATTCATGTATTGGCG--- 1077
QY 416 LysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeu--- 434
Db 1078 -----TCATTATTGAAACGATGCTGATTACCGACCTCTAAGATTATCA 1122
QY 435 ---PheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAsp 453
Db 1123 CAAGTGATTGAGCAAACTGGTTAGCCCGCAGACATGGCTTGNATGTTATTG---GAT 1179
QY 454 GluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIle----- 468
Db 1180 CGTACATTATTGCGGATATTAAAGCTGTATAGTAAACACAGCACTATTGCTAGAGACAT 1239
QY 469 -----HisSerThrLeuThrLeu-----Asp 475
Db 1240 GTACAAAATTCATTAATTAATGAATTAATAAAGATGGCATCCCGACGCTATGTTGTGAGAGAT 1299
QY 476 AsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeu 495
```

```
Db 1300 CTTGAAGGCATTTGTCTATCTAGAACGATTGCTACTGAAAAACAATCTGTGCCAAAT--- 1356
QY 496 AspValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeu 515
Db 1357 ---GTTGTGGCAGCATCTACGCCCT-----GTTGTATATGCACATGATGAAGCG 1401
QY 516 PheIleAsnLysLysTyrPheLeuValIleAspArg-----AlaIle 529
Db 1402 TGGCATCGCTTAAATATTACGTTGTAGTAAATCACTTAGGACATTAGTATCACTATT 1461
QY 530 GlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeu----- 542
Db 1462 GGTAAAGCACAAGATGAAGTTGTGTATGGCACTTGTAGCGCATCGTCTTATGACT 1521
QY 543 -----LysGlu---AspSerAsnProValPheAspLysThrLysAsnArgValTyrThr 559
Db 1522 TGGAAAAAAGAAATACGCAATACGCGATATTGTTGACTGTGTAGAAAGATTTATATCAA 1581
QY 560 Thr 560
Db 1582 ACG 1584

RESULT 14
US-09-815-242-8108
; Sequence 8108 Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8108
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1737)
US-09-815-242-8108

Alignment Scores:
Pred. No.: 0.0381 Length: 1737
Score: 110.50 Matches: 105
Percent Similarity: 34.25% Conservative: 94
Best Local Similarity: 18.07% Mismatches: 211
Query Match: 3.16% Indels: 171
DB: 10 Gaps: 26
```

US-09-802-285A-2 (1-659) x US-09-815-242-8108 (1-1737)

```
QY 91 IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis 110
Db 43 ATCGAACGGATTTAAATGAATATTTAGAGAGAAAATTTATATCAAGTACCACCTCAA 102
QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
Db 103 AAT-----CATCAATGGCTATACAA 123
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpGlnAlaMetAla---Leu 149
Db 124 TTATCTGAGCTCGAAACGTTAAACAGGTCAATCGCTTATGGTCTCGGATGGGCATCAT 183
QY 150 ValTyrHisAlaThrGly-----AspGluLysTyrAlaArgGluTrpValTyrGln 166
Db 184 ATGTATCATTCAGAGTATGGCTTATGATGGTAAAGTAAAGAAATTAACAACATATAAA 243
QY 167 TyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheVal 186
Db 244 GAAGCCATGGCGGATTTTGCACATATGGCTCAAAGTGCAGATPAATCAACGGCAGTG 303
QY 187 TrpArgProLeuGluValSerAspArgValGlnSerLeuProProThrPheSerLeuPhe 206
Db 304 CAACAACATATG----- 315
QY 207 ValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGln 226
Db 316 -----CGCAAAATTTATCTGACATCATATAGCATTCATCGC 354
QY 227 GlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGluAla 246
Db 355 ACGGCGGTTATTTGCMAAGTAAACAAATAGACTACGACAGATCGCTATATCGTTTCA 414
QY 247 GlnArgAsnLeuPheAlaGlyValSerPhe---ProGluPheLysAspSerProArgTrp 265
Db 415 GAACAATCTTTATCTTAGTCTATCCATTTTCATCGACTCCTTAAGAGTGAAGTGGGTT 474
QY 266 ArgGlnThrGlyLeuSer----- 271
Db 475 TCAGAAGCAGATTTAGAGAAATATGACCCGAAATGTCATACATTCACATTCGAAATG 534
QY 272 ---ValLeuAsnThrGluIleLysLysGlnValTyrAlaAspGlyMetGlnPheGluLeu 290
Db 535 TTAGCTGTGTCATCAAGATCTTCTGCTCACGCGCTATGTAGAGGTAAAGAGATCAGGTT 594
QY 291 SerProIle---TyrHisValAlaAlaIleAspIle-----PheLeu 303
Db 595 GAGAAAGTGTGTATCAATATAGCAGATATAGACATATCAGAGATACCTTAAAGATTTTAT 654
QY 304 LysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrVal 323
Db 655 TTATTACCAATACATCCTTATCAATCAATGTTGTCACAGCATCCACAG---TATTTG 711
QY 324 GlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnThr 343
Db 712 CAATATAGTGAACAAGGTTTAAATAAAAGACCTTGGCGTTTC----- 753
QY 344 ProMetPheGlyAspSer----- 349
Db 754 -----GGTGATTCAGTTTACCAGCGTCTTCGGTTAGAACTGTATTTTCAAAGCA 804
QY 350 -----TrpIleThrAspLysAsnPhe---ArgMetAla 359
Db 805 TTAAACATTTATTTAAATACCGATACAGTGTAAATTTACTAAATTTATACGTACGAT 864
QY 360 GlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIle----- 374
Db 865 GATCTTGAACAGATTGAACGGCAATCGATGTCGACAAAGTTATCGCATCAGTCAAAGAT 924
QY 375 -----LysTyrPheAlaThrAspGlyLysGlnGlyLysAlaPro 387
Db 925 ----- 397
```

```
Db 925 GAGGTTGAAAACACCCCATTTTAAATGTATGTTTGAAGAAGGATATCGTGCATTTTACCG 984
QY 388 AsnPheLeuSerLysAla-----LeuSerAsnAlaGlyPheTyr 400
Db 985 AATCCATTAGGGCAACACAGCTGAACCTGAAATGGATTTATTAACAATAGTGCATGATT 1044
QY 401 ThrPheArgSerGly-----TrpAspLysAsnAlaThrValMetValLeu 415
Db 1045 GTT---CGTGAAGGATACCGAATTAACCATGCTGATGAAGATATTCATGTATTGGCG--- 1098
QY 416 LysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeu--- 434
Db 1099 -----TCATTATTGAAACAGATGCTGATTACCGACCTCTAAGTTATCA 1143
QY 435 ---PheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAsp 453
Db 1144 CAAGTGATTGACAAAGTGGTTTAGCGCCAGAGCATGGCTTGAATGTTATTGTTG---GAT 1200
QY 454 GluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIle----- 468
Db 1201 CGTACATATTGCCGATATTAAGCTGTATAGTAACACAGGCATTAGCTCTAGAAGCACAT 1260
QY 469 ---HisSerThrLeuThrLeu-----Asp 475
Db 1261 GTACAAAATTCATTAAATGAATTAAGAGATGCATACCCGACGTATGCTTTGTCAGAGAT 1320
QY 476 AsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnLeu 495
Db 1321 CTTGAAGGCATTTGTCTATCTAGAACGATGCTACTGAAACAAACACTTGTGCCAAAT--- 1377
QY 496 AspValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeu 515
Db 1378 ---GTTGGCGCAGCATCTAGCCCT-----GTTGTATATGCACATGATGAAGCG 1422
QY 516 PheIleAsnLysLysTyrPheLeuValIleAspArg-----AlaIle 529
Db 1423 TGGCATCGCTCTAAATATATACGTTGTAGTAAATCACTTAGGACATTTAGTATCAACTATT 1482
QY 530 GlyGluAlaThrGlyAsnLeuGlyValHisTyrGlnLeu----- 542
Db 1483 GGTAAAGGCACAAAGATGAAGTTGTGTATGGCAACTGTAGCCGATCGTCTTATGACT 1542
QY 543 -----LysGlu---AspSerAsnProValPheAspLysThrLysAsnArgValTyrThr 559
Db 1543 TGGAAAAAAGATACACGCGAATACCCAGTATTTGTTGACTGTGTAGACATTTATATCAAA 1602
QY 560 Thr 560
Db 1603 ACG 1605
```

RESULT 15

US-09-070-927A-91/C

; Sequence 91, Application US/09070927A

; Patent No. US20020120116A1

; GENERAL INFORMATION:

APPLICANT: Charles A. Kunsch

Patrick J. Dillon

Steven Barash

; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 982

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII TEXT

; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070,927A
 FILING DATE: 04-May-2000
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,655

FILING DATE: 1997-05-16

APPLICATION NUMBER: 60/044,031

FILING DATE: 1997-05-06

APPLICATION NUMBER: 60/066,009

FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB369

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 91:

SEQUENCE CHARACTERISTICS:

LENGTH: 10194 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 91:

US-09-070-927A-91

Alignment Scores:

Pred. No.:	0.596	Length:	10194
Score:	110.50	Matches:	143
Percent Similarity:	36.54%	Conservative:	115
Best Local Similarity:	20.25%	Mismatches:	274
Query Match:	3.16%	Indels:	178
DB:	11	Gaps:	32

US-09-802-285A-2 (1-659) x US-09-070-927A-91 (1-10194)

QY	15	IleAlaLeuSerSerGlyAsnIleLeuAlaGlnSerSerSer	-----	28
Db	9401	GTATCTCTTAAAGGAGGAGATATCTTCACATTCGAAACTGCCGACAAAGATTTTGTGT	9342	
QY	29	-----IleThrArgLysAspPheAspHisIle	-----	AsnLeuGlu 40
Db	9341	AGATTTTAAATGAAAAAGATTTTCAAGAATATATGAAAAATCCAGTATCMAGTACTGAA	9282	
QY	41	TyrSerGlyLeuGluLysValAlaLysAlaValAlaAlaGlyAsnTyrAspAspAlaAla	60	
Db	9281	ACTAGCAACAAACAAACAGTAAACTCTGATGTTTCTAAAGTAGTAGCCCAAGATAATAA	9222	
QY	61	LysAlaLeuLeuAlaTyrTyrArgGlyLysSerLysAlaArgGluProAspPheSerAsn	80	
Db	9221	CAATCTGATGATCTGAAAAAAGAGTAAAGTACTGACCGAGTCTGATGTAGTACT	9162	
QY	81	AlaGluLysProAla	-----	AspIleArgGlnProIleAspLysVal 94
Db	9161	AATACTTTTACCGAGCAAGATAAAATCTAATGACATTAATAAGTAGTAGCAGATGAGCCA	9102	
QY	95	ThrArgGlu	-----	-MetalaAspLysAlaLeuValHisGlnPheGln- 108
Db	9101	ACCTTAGAACAAACACCGCTTAGATACCTTTAGTAGTACATCAATTTAATGATATGTAT	9042	
QY	109	ProHisLysGlyTyrGlyTyrPheAspTyrGly-LysAspIleAsnTyrGlnMetTrpPr	128	
Db	9041	CTTTATAAGGAGCAAAATG-----CATTCATATTCGGCGTCATCCACCATGGAC	8988	
QY	128	oValLysAspAsnGluValArgTrpGlnLeuHisArgValLysTyrTrpGlnAlaMet	147	
Db	8987	GCAAAAAGATGTT-----AAATGGTATCAAAAAGTAGT	8955	
QY	148	-AlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpValTyr	165	
Db	8954	TGCTACTATAGTAAATGCATATGAGCAAAAGAGAGAGCTAATG--TGGAAATACACATA	8897	
QY	166	-----GlnTyrSerAspTrp	170	

Db	8896	ACTCCAAAGTCAGATAGCGAGCTTGTAGAAATTTATGATTTATTAATCTTCTATT	8837
QY	171	-----AlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTr	187
Db	8836	ATTTTATTAAAGAGAGAAACCTTCGGGTTTTTTCTTTTATCACAACCAACGACATA	8777
QY	187	pArgProLeuGluValSerAspArgValGln-SerLeuProProThrPheSerLeuPheV	207
Db	8776	TATTCGTATTTCGTATTATGAT--ATACGATACTTGATCTTACTTATAACATATTC-	8721
QY	207	AlaAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnG	227
Db	8720	--ATAAAAAATCGTCTTAAATGTTAAATTA	8688
QY	227	lnAlaAspTyrLeuSerThrHisTyrAlaGlnGlnGlyAsnHisArgLeuPheGluAlaG	247
Db	8687	-----AAAAATCCGAGAGGTTGGTATTATTCGCAG	8657
QY	247	lnArgAsnLeuPheAlaGlyValSerPhePro-----GluPheLysAspSer	262
Db	8656	AACCAAGAAATTAGCTAATGGAATGCGAGTTAAAGATTAAATATAAAGACTCTATTA	8597
QY	263	--ProArgTrpArgGlnThrGlyLysSer-----ValLeuA	274
Db	8596	CAATCAATGGAATAATAATAATCAATCTACTAGATCTACTAAGAGGCTGTAGAGATGCTG	8537
QY	274	snThrGluLeuLysLysGlnValTyr-----AlaAspGlyMetGlnPheGluLeuSerP	292
Db	8536	AAACAGAGTTTAAATCAAAAATATGCGTGGTGAACAACTGAAGCAATTTAAACTTTTAG	8477
QY	292	roIleTyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgV	312
Db	8476	ATTTTATGATATT--TGGTTGATACATTTTAAAGAATAAAGTGTCTGCTGGCGGT-	8421
QY	312	AlaAsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetA	332
Db	8420	-----ATGCAGAAAA	8408
QY	332	laLeuIleSerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleT	352
Db	8407	CITTAACAAGAAAGATTTTACATGATTC-----TTTGGGACAAACAATTCG	8360
QY	352	hrAspLysAsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnG	372
Db	8359	TA---AAAGCGCTCACCAAGTCAAAATATCAGCAATG	8325
QY	372	lnAlaIleLysTyrPheAlaThrAspGly-----LysGlnGlyLysAlaProAsnP	389
Db	8324	-----ATTAACCTGGCTTGCAAAACCTGGCAATATCAATGAAAGAGCTAGCAGTTGAG	8270
QY	389	heLeuSerLysAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTyrAspLysA	409
Db	8269	CGGTGAGCAACAGACATAA-CATTGTAAATCTATCTTTTAAAGCGT-TGCACATGCA	8212
QY	409	snAlaThrValMetValLeuLysAla-----SerProGlyGluPheHisAlaGlnP	427
Db	8211	ATATATTCAATCAATCCAAACCAAGAAATATAAGTTA-ACTGGCCCAAGTACCAAGAAAG	8153
QY	427	roAspAsnGlyThrPhe-----GluLeuPheIleLysG	438
Db	8152	AATCAATATAAACCATATCTTATGATGATTTAACTAAATTTAAAGAGCTTTATATCTC	8093
QY	438	lyArgAsnPheThrProAspAlaGlyValPheVal-----TyrSerGlyAspGluAlaI	456
Db	8092	GTGAAACACCAACAGTAAATATCTCTTTTAGTACAACTATATACTGGAGCTCGT---	8037
QY	456	leMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspA	476
Db	8036	-----TATCAAGAGGTTTCTGCGTTGACTTGGGATAGCATTCATGATG	7997
QY	476	snGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuA	496

Db 7996 AAAGTAACGAGATTATTAAAGATAAAAAATGCCTATAAGTATGATGCTGGCCAGTATAGAA 7937
QY 496 spValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuP 516
Db 7936 TTGGCCCTACAAAATCAGAACGAGGTTCGAAATATTGATGCCCTTCTCTTTCT 7877
QY 516 heileAsnLysLysTyrPheLeuValIleAspArgAlaIle-----GlyGluAlaThrG 534
Db 7876 ATTATTAAAAAAATGAATGGACAAAAGAAAAATATTGACAGGAGATTACGTA 7817
QY 534 lyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAspLysThrL 554
Db 7816 ATCCCCAAAATTTTGCTTTTGTAGTAAAAAAATTCATGGCCCATCTCAAAATTCATCTG 7757
QY 554 ysAsnArgValTyrThr-----T 560
Db 7756 TGAAATAATATATAAAGAAACATGTGAACCTAGCAAAATATTGAAAGAAATTCATCTCATA 7697
QY 560 hrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAsp-----A 576
Db 7696 GTTTTAGACATGCCCGTTCTGACTTATTAACTAGCTGAAGCTGATCCAATATATATCA 7637
QY 576 rgThrSerLeuAsnGluGluGlyLysValSerTyrValTyr-AsnLysGluLeuLys 595
Db 7636 AAACACAGCTAGCACATAAAGAGATAATACAGCTCATATGAATACGCATCAGCTACTGAAG 7577
QY 596 ArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSer 615
Db 7576 CAA-----ATAGAAAAAGACAAAGAAAAATATGAAAAACAAATACAAAGATATTT 7526
QY 616 ileValTyrPro 619
Db 7525 TGTAATAATACCCA 7514

Search completed: August 4, 2003, 13:41:05
Job time : 4172 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2003, 10:13:19 ; Search time 85 Seconds
(without alignments)
2377.647 Million cell updates/sec

Title: US-09-802-285A-2

Perfect score: 3494

Sequence: 1 MTKIFKRIIVFAVIALSSG.....KGLNLTITNGKQQLVLP 659

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US0802285/runat_28072003_190513_9740/app.query.fasta_1.839
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0802285.@cgn_1_1_40 -runat_28072003_190513_9740 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3494	100.0	1980	1	US-08-258-639A-3
2	3494	100.0	1980	2	US-08-900-951-3
3	3494	100.0	1980	5	PCT-US95-07391A-3
C	176	5.0	3763	4	US-08-961-527-186
5	132.5	3.8	2097	4	US-09-134-001C-795
6	123.5	3.5	3414	1	US-07-973-320-1
7	123	3.5	1542	1	US-08-328-962-1
C	120.5	3.4	4403765	4	US-09-103-840A-2
C	120.5	3.4	4411529	4	US-09-103-840A-1
10	115.5	3.3	3414	1	US-07-973-320-3
11	115.5	3.3	3504	1	US-08-620-717A-8
12	115	3.3	2859	5	PCT-US96-05320A-637

13	115	3.3	16535	4	US-08-961-527-74	Sequence 74, Appl
14	114.5	3.3	3504	1	US-08-485-568A-5	Sequence 5, Appl
15	114.5	3.3	3504	2	US-08-590-554A-5	Sequence 5, Appl
16	114.5	3.3	3504	2	US-09-184-223-5	Sequence 5, Appl
17	113.5	3.2	4765	1	US-08-750-532-8	Sequence 8, Appl
18	113.5	3.2	4765	4	US-08-894-818B-7	Sequence 7, Appl
19	113.5	3.2	4765	4	US-09-445-472-5	Sequence 5, Appl
20	112.5	3.2	2339	1	US-08-258-639A-1	Sequence 1, Appl
21	112.5	3.2	2339	2	US-08-900-951-1	Sequence 1, Appl
22	112.5	3.2	2339	5	PCT-US95-07391A-1	Sequence 1, Appl
23	111	3.2	2784	4	US-09-134-001C-1994	Sequence 1994, Ap
24	109	3.1	2858	3	US-08-816-346-3	Sequence 3, Appl
25	109	3.1	2858	3	US-08-816-346-57	Sequence 57, Appl
26	109	3.1	2858	3	US-09-335-411-3	Sequence 3, Appl
27	109	3.1	2858	3	US-09-335-411-57	Sequence 57, Appl
28	109	3.1	2859	2	US-08-788-674-7	Sequence 7, Appl
29	109	3.1	34303	2	US-08-735-609-4	Sequence 4, Appl
30	109	3.1	34303	2	US-08-735-609-4	Sequence 4, Appl
31	109	3.1	34303	3	US-09-315-372-4	Sequence 4, Appl
32	109	3.1	34303	3	US-09-244-752-4	Sequence 4, Appl
33	109	3.1	34303	3	US-09-245-497-4	Sequence 4, Appl
34	109	3.1	34303	4	US-09-562-919-4	Sequence 4, Appl
35	109	3.1	34382	2	US-08-374-483-6	Sequence 6, Appl
36	109	3.1	35408	4	US-08-973-334-3	Sequence 3, Appl
37	109	3.1	35408	4	US-09-563-869A-3	Sequence 3, Appl
38	109	3.1	35408	4	US-08-549-489-3	Sequence 3, Appl
39	109	3.1	35935	2	US-08-735-609-1	Sequence 1, Appl
40	109	3.1	35935	2	US-08-735-609-1	Sequence 1, Appl
41	109	3.1	35935	3	US-08-379-452-43	Sequence 43, Appl
42	109	3.1	35935	3	US-09-315-372-1	Sequence 1, Appl
43	109	3.1	35935	3	US-09-244-752-1	Sequence 1, Appl
44	109	3.1	35935	3	US-09-245-497-1	Sequence 1, Appl
45	109	3.1	35935	3	US-09-409-670-43	Sequence 43, Appl

ALIGNMENTS

RESULT 1

US-08-258-639A-3
; Sequence 3, Application US/08258639A
; Patent No. 5681733

; GENERAL INFORMATION:
; APPLICANT: Su, Hongsheng

; APPLICANT: Blain, Francoise

; APPLICANT: Bennett, Clark

; APPLICANT: Gu, Kangfu

; APPLICANT: Zimmermann, Joseph

; APPLICANT: Musil, Roy

; TITLE OF INVENTION: Nucleic Acid Sequences And Expression

; TITLE OF INVENTION: Systems For Heparinase II And Heparinase III Derived From

; TITLE OF INVENTION: Flavobacterium heparinum

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hale and Dorr

; STREET: 1455 Pennsylvania Avenue, N.W.

; CITY: Washington, D.C.

; COUNTRY: U.S.A.

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/258,639A

; FILING DATE: 10 JUNE 1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Healey, William J.

; REGISTRATION NUMBER: 36,160

; REFERENCE/DOCKET NUMBER: 104385.116

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)942-8400

Db 1861 GACGGCCAGAGGCTCCAGAGATCAGCATACGGGAAACAAAGGCAATGATTGTGAGAA 1920
QY 641 GlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnGlnLeuValLeuValPro 659
Db 1921 GGCAGCTTAATCTACCCCTTACCATTAACGGAACAAACAGCTTGTTGGTTCT 1977
RESULT 2
US-08-900-951-3
Sequence 3, Application US/08900951
Patent No. 5919693
GENERAL INFORMATION:
APPLICANT: Su, Hongsheng
APPLICANT: Blain, Francoise
APPLICANT: Bennett, Clark
APPLICANT: Gu, Kangfu
APPLICANT: Zimmermann, Joseph
APPLICANT: Musil, Roy
TITLE OF INVENTION: Nucleic Acid Sequences And Expression
TITLE OF INVENTION: Systems For Heparinase II And Heparinase III Derived From
TITLE OF INVENTION: Flavobacterium heparinum
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRES:
ADDRESSEE: Hale and Dorr
STREET: 1455 Pennsylvania Avenue, N.W.
CITY: Washington, D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS/MS DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900,951
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/258,639
FILING DATE: 10 JUNE 1994
ATTORNEY/AGENT INFORMATION:
NAME: Healey, William J.
REGISTRATION NUMBER: 36,160
REFERENCE/DOCKET NUMBER: 104385.116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)942 8400
TELEFAX: (202)942 8484
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1980 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-900-951-3
Alignment Scores:
Pred. No.: 0 Length: 1980
Score: 3494.00 Matches: 659
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0
US-09-802-285A-2 (1-659) x US-08-900-951-3 (1-1980)
QY 1 MetThrThyIysIlePheLysArgIleIleValPheAlaValIleAlaLeuSerSerGly 20
Db 1 ATGACTACGAAATTTTAAAGGATCATGTATTGCTGTGTAATTGCCCTATCGTCGGA 60
QY 21 AsnIleuAlaGlnSerSerIleThrArgLysAspPheAspHisIleAsnLeuGlu 40
Db 61 AATATACTTGCAAAAGCTCTCCATTACGAGGAAAGATTTTGACCACATCAACCTTGAG 120

QY 41 TyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAspAlaAla 60
Db 121 TATTCGGAGCTGAAAAGGTTAATAAGCAGTTGTCTCCGCGCAACTATGACGATGCGGCC 180
QY 61 LysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsn 80
Db 181 AAAGCATTTACTGGCATACTACAGGAAAAAAGTAAGCCAGGGAACCTGATTTTCAGTAAT 240
QY 81 AlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAsp 100
Db 241 GCAGAAAAGCCTCGCATATATACGCCAGCCCATAGATGAAGGTTACGGCTGAATGGCCGAC 300
QY 101 LysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLys 120
Db 301 AAGGCTTTGGTCCACAGTTTCAACCCGACAAAGGCTACGGCTATTTTGATTTATGGTAAA 360
QY 121 AspIleAsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArg 140
Db 361 GACATCAACTGGCAGATGTGGCCGTTAAAGACAAATGAAGTACGCTGGCAGTTGCACCGT 420
QY 141 ValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla 160
Db 421 GTAAATATGGTGGCAGCTATGGCCCTGGTTTATCACGCTACGGGCGATGAAAAATATGCA 480
QY 161 ArgGluTrpValTyrGlnTyrSerAspTyrAlaArgLysAsnProLeuGlyLeuSerGln 180
Db 481 AGAAGATGGGTATATCAGTACAGCATTTGGCCGAGAAAACCCATTGGCCCTGTGCGAG 540
QY 181 AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro 200
Db 541 GATAATGATAAAATTTGTGTGGCGGCCCTTGAAGTGTGCGACAGGGTACAAAGTCTTCCC 600
QY 201 ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220
Db 601 CCAACCTTTCAGCTTATTTGTAACTCCGAGCCCTTTACCCGACCTTTTAAATGGAATTT 660
QY 221 LeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsn 240
Db 661 TTAAACAGTTACCAACAGCCGCGATTATTAATCTACGCATTATGCCGAACAGGGAAC 720
QY 241 HisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys 260
Db 721 CACCGTTTATTTGAAGCCCAACCAAGCACTGTTGCGAGGGTATCTTTCCTGATTTAAA 780
QY 261 AspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleLysLysGln 280
Db 781 GATTCCACCAAGATGGAGGCAAAACCGGCATATCGTGTCTGAACACCGAGATCAAAAAACAG 840
QY 281 ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAsp 300
Db 841 GTTTATGCCGATGGGATGAGTTTGAACTTTTCCCAATTTACCATGTAGCTGCCATCGAT 900
QY 301 IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln 320
Db 901 ATCTTCTTAAAGGCCCTATGGTTCTGCAAAACAGTTTAACCTTGAAAAAGAAATTTCCGCAA 960
QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
Db 961 TCTTATGTACAAACTGTAGAAAATATGATTATGGCGCTGATCAGTATTTTCACTGCCAGAT 1020
QY 341 TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGln 360
Db 1021 TATACACCCCTATGTTTGGAGATTTCAGATTTACAGATAAAATTTTCAGGATGGCACAG 1080
QY 361 PheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAsp 380
Db 1081 TTTTGGCAGCTGGCCCGGGTTTTTCCGCGCAACACCGCCATTAATATTTTGTACAGAT 1140
QY 381 GlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyr 400
Db 1141 GGCAAAACAAAGGTAAGCGCGCTAACTTTTATCCAAAGCATTTGAGCAATTCAGGCTTTTAT 1200
QY 401 ThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProPro 420

Db 1201 AGTTTAGAGCGGATGGATATAAATGCAACCGTTATGTTATTAAGACCGAGTCTCTCC 1260
QY 421 GlyGluPheHisAlaGlnProAspGlyThrPheGluLeuPheIleLysGlyArgAsn 440
Db 1261 GGGGAATTTCTGCGCCAGCGGATACAGGCACTTTTGAACCTTTTATAAAGGCGAGAAAC 1320
QY 441 PheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArg 460
Db 1321 TTTACCCAGAGCGCGGGTATTTGTTATAGCGGACGAGACCATCATGAACCTGGG 1380
QY 461 AsnTyrPyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnMetVal 480
Db 1381 AACTGGTACCGTCAACCCGATACACAGCAGCTTACACTCGACATCAAAATATGGTC 1440
QY 481 IleThrLysAlaArgGlnAsnLysTyrGluThrGlyAsnLeuAspValLeuThrTyr 500
Db 1441 ATTACCAAGCCCGCAAAACAAATGGGAAACAGGAAATAACCTTGCTGTACCTAT 1500
QY 501 ThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLys 520
Db 1501 ACAACCCAGCTATCCGAACTTGGACCATACGCGAGTGTACTTTTCATCAACAAAAA 1560
QY 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTyr 540
Db 1561 TACTTTCTGTCATCGATAGGCAATAGCGAGCTACCGAAACCTGGCGGTACACTGG 1620
QY 541 GlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThr 560
Db 1621 CAGCTTAAAGAGACAGCAACCTGTTTTCGATAGACAAAGAACCGGGTTTACACCACT 1680
QY 561 TyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn 580
Db 1681 TACAGAGATGGTAACCACTGATGATCCCAATCGTTGAATCGGACAGGACCACTCAAT 1740
QY 581 GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal 600
Db 1741 GAAGAAGAGGAAAGGTACTTATGTTTACATAAGGAGCTGAAAGACCTGCTTTCGTA 1800
QY 601 PheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyr 620
Db 1801 TTTGAAAGCCTTAAAGAAATCGCGCACACAAAAATTTTGTAGTATAGTTTATCCATAC 1860
QY 621 AspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLys 640
Db 1861 GACGGCCAGAGCTCCAGATTCAGATACGATACGCGGAAACAGGGCAATGATTTTGAGAA 1920
QY 641 GlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnGlnLeuValLeuValPro 659
Db 1921 GCGAAGCTTAATCTAACCTTACCATTAACGGAAACAAACAGCTTGTGTGTTCT 1977

RESULT 3

PCT-US95-07391A-3
; Sequence 3, Application PC/TUS9507391A
; GENERAL INFORMATION:
; APPLICANT: IBEX TECHNOLOGIES and
; APPLICANT: ZIMMERMANN, Joseph
; TITLE OF INVENTION: Nucleic Acid Sequences And Expression
; TITLE OF INVENTION: Systems For Heparinase II And Heparinase III Derived From
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hale and Dorr
; STREET: 1455 Pennsylvania Avenue, N.W.
; CITY: Washington, D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/07391A
; FILING DATE: 09-JUNE-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/258,639
; FILING DATE: 10 JUNE 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Hollie L.
; REGISTRATION NUMBER: 31,321
; REFERENCE/DOCKET NUMBER: 104385.116PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)942-8400
; TELEFAX: (202)942-8484
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1980 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US95-07391A-3

Alignment Scores:

Pred. No.: 0 Length: 1980
Score: 3494.00 Matches: 659
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-09-802-285A-2 (1-659) x PCT-US95-07391A-3 (1-1980)

QY 1 MetThrThrLysIlePheLysArgIleIleValPheAlaValIleAlaLeuSerSerGly 20
Db 1 ATGACTACGAAATTTTAAAGAGATCATGTTATTTGCTGTATTCCTATTCGTCGGGA 60
QY 21 AsnIleLeuAlaGlnSerSerSerIleThrArgLysAspPheAspHisIleAsnLeuGlu 40
Db 61 NATATCTTGCAACAAGCTCTTCATTACCAGGAAGATTTTGACCACATCAACCTTGAG 120
QY 41 TyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAlaAla 60
Db 121 TATTCGGACTGGAAAAAGGTTAATAAAGCAGTTGCTCGCGCAACTATCAGCATGGGCC 180
QY 61 LysAlaLeuLeuAlaTyrThrArgGluLysSerLysAlaArgGluProAspPheSerAsn 80
Db 181 AAAGCATTTACTGGCATACTACAGGAAAAAAGTAGAGCCAGGGAACCTGATTTTCAGTAT 240
QY 81 AlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAsp 100
Db 241 GCAGAAAAGCCTGCCGATATACGCCAGCCCATAGATAAGTTACGGTGAATGGCCGAC 300
QY 101 LysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLys 120
Db 301 AAGGCTTTGGTCCACAGTTTCAACCGCAACAAGGCTACGGCTATTTTGTATATGTATAA 360
QY 121 AspIleAsnTyrGlnMetTyrProValLysAspAsnGluValArgTyrGlnLeuHisArg 140
Db 361 GACATCACTGGCAGATGTGGCGGTAAAGACATCAAGTAGTACCTGGCAGTTGCCACCT 420
QY 141 ValLysTyrTyrGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla 160
Db 421 GTAAAAATGGTGGCAGGCTATGCGCTGTTTATCAGCTACGGCGATGAAAAATATGCA 480
QY 161 ArgGluTyrValTyrGlnTyrSerAspTyrAlaArgLysAsnProLeuGlyLeuSerGln 180
Db 481 AGAGAATGGGTATATCAGTACAGCGATTGGGCCAGAAAAAACCATTGGGCTGTGCGCAG 540
QY 181 AspAsnAspLysPheValTyrArgProLeuGluValSerAspArgValGlnSerLeuPro 200
Db 541 GATAATGATAAATTTGTGGCGGCCCTTGAAGTGTGGACAGGAGGTACAAAGCTCTCCC 600
QY 201 ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220

Db 601 CCAACCTTCAGCTATTATTTGTAACCTCCGAGCCTTTACCCAGCCTTTTAAAGGAAATTT 660
QY 221 LeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsn 240
Db 661 TTAACAGATTACCAACCAACAGGCGGATTTATTTATCTACGCAATATATGCGGAACAGGGAAC 720
QY 241 HisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys 260
Db 721 CACGGTTATTTAGGCCCAACCGCACTTGTTTGCAGGGGTATCTTTCCCTGAAATTTAAA 780
QY 261 AspSerProArgTyrArgGlnThrGlyIleSerValLeuAsnThrGluIleLysLysGln 280
Db 781 GATTCAACCAAGATGGAGGCAACCGGCATATCGGTGCTGAACACCGAGATCAAAAACAG 840
QY 281 ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaIleAsp 300
Db 841 GTTTATGCCGATGGATGCAGTTTGAACTTTCACCAATTTACCAATTTAGCTGCGCATCGAT 900
QY 301 IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln 320
Db 901 ATCTTCTTAAGCCCTATGTTTCTGCAAAACGAGTTAACCTTGAAAAGAAATTTCCGCAA 960
QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
Db 961 TCTTATGTACAACTGTAGAAATATGATATATGCGCTGATCAGTATTTCACTGCCAGAT 1020
QY 341 TyrAsnThrProMetPheGlyAspSerTyrIleThrAspLysAsnPheArgMetAlaGln 360
Db 1021 TATAACACCCCTATGTTTGGAGATTCATGATTTACAGATAAAATTTTCAGGATGGCACAG 1080
QY 361 PheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAsp 380
Db 1081 TTTGCCAGCTGGCCCGGGTTTCCCGGCAACAGGCCCAATAAATTTTGCTACAGAT 1140
QY 381 GlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyr 400
Db 1141 GGCAAAACAGGTAAAGCGCCTAACTTTTATCCAAAGCATTTAGCAATGCAGGCTTTAT 1200
QY 401 ThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProPro 420
Db 1201 ACCTTTAGAGCGGATGGGATAAAATGCAACCGTTATGTTATTAAGCGCAGTCCTCCC 1260
QY 421 GlyGluPheHisAlaGlnProAsnGlyThrPheGluLeuPheIleLysGlyArgAsn 440
Db 1261 GGGGAATTCATGCCACCGCGGATACGGGACITTTGAACTTTTATAAAGGGCAGAAAC 1320
QY 441 PheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArg 460
Db 1321 TTTACCCACAGCCCGGGGTATTTGTATAGCGGCGAGCAACCATCATGAACCTCGCG 1380
QY 461 AsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetVal 480
Db 1381 AACTGTGTACCGTCAAAACCCGCATACACAGCAGCGTTACATCGACATCAAAATATGGTC 1440
QY 481 IleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyr 500
Db 1441 ATTACCAAGCCCGCAAAACAATGGAAACAGAAATAACCTTGATGGCTTACCTAT 1500
QY 501 ThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLys 520
Db 1501 ACCAAACCAAGCTATCGAATCTGGACCATCAGCGAGTGTACTTTTCATCAACAAAAA 1560
QY 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrp 540
Db 1561 TACTTTCTGTGTCATCGATAGGCAATAGCGGAACCTACCGGAAACCTGGCGGTACACTGG 1620
QY 541 GlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrThr 560
Db 1621 CAGCTTAAAGAGACAGCAACCTGTTTTCGATTAAGCAAAAGAACCGGGTTTACACCACT 1680
QY 561 TyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn 580

Db 1681 TACAGAGATGGTTAAACACCTGATGATCCCAATCGTTGAATCGGACAGACGACCTCAAT 1740
QY 581 GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal 600
Db 1741 GAAGAAGAAGAAAGGTATCTTATGTTTACAATAAGAGAGCTGAAAAGACCTGCTTTCGTA 1800
QY 601 PheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyr 620
Db 1801 TTTGAAAAGCCTAAAAGAAATGCCGCACACAAAATTTTGTCAATATAGTTTATCCATAC 1860
QY 621 AspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLys 640
Db 1861 GACGGCCAGAGGCTCCAGAGATCAGATACGGAACAAAGGCAATGATTTTGAGAAA 1920
QY 641 GlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnGlnLeuValLeuValPro 659
Db 1921 GGCAAGCTTAATCAACCTTACCATTAACGGAACAAACAGCTTGTGTTGGTTCT 1977

RESULT 4

US-08-961-527-186/c
; Sequence 186, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 186:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3763 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-186

Alignment Scores:

Pred. No.: 5,63e-10 Length: 3763
Score: 176.00 Matches: 116
Percent Similarity: 34.63% Conservative: 80
Best Local Similarity: 20.49% Mismatches: 231
Query Match: 5.04% Indels: 140
DB: 4 Gaps: 20

US-09-802-285A-2 (1-659) x US-08-961-527-186 (1-3763)

QY 71 SerLysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnPro 90

Db 2173 TCMAAGACTATGAGAAAGTCAAGATTCCTTGAACGCTTGATGATGCAATCGTTTATG 2114

```
QY 91 IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis 110
Db 2113 TTGTAGT---CCTGGATATGGAGCTTGTCAAAATCCATCAATCCAGCG--- 2060
QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
Db 2059 -----ATGGTATGGGATCAAGTAATTTGAAGAT 2033
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpGlnAlaMetAlaLeuVal 150
Db 2032 GATCCAGATGGCTTATATGCTCAATCAGCAAGATATCTCTGCAGCTTTATGATAGGG 1973
QY 151 TyrHisAlaThrGlyAspGluLysTyrAlaArg-----GluTrp 163
Db 1972 TATCTGGTAGAGGATAGGACTATATCAAAAGTCAAGTCTCTTCTATTGATGG 1913
QY 164 ValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAsp 183
Db 1912 ATTGACGAGTGAGAGAAATTTCTCTCAATCCTTGATGACTAGAACCTTGGATACGGGT 1853
QY 184 -----LysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProPro 201
Db 1852 ATTGCTCTCTTACTTGGTTGAAACTA----- 1826
QY 202 ThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeu 221
Db 1825 ---CCTTCTCTCTTGAATTTGACTTGTAGAGGAGAAAGAACTAGAGAAATTTG 1769
QY 222 AsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln----- 238
Db 1768 GTCAGTCTAGAAAAGCAGATTGACTTTATGAAAAGCTACTATCGCGCCAACTACACCTT 1709
QY 239 GlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGlu 258
Db 1708 AGTAACCTGGGGATTTTACAAACAATCCGATGCTTGCATCTATCATTTCTTTTCAGAT 1649
QY 259 PheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluLeuLys 278
Db 1648 AAGATGATCTAGAGAAGCTTACCATTTTGCTTCAGAGAGTTGAAACAGCAATTTAG 1589
QY 279 LysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisVal----- 296
Db 1588 ACACAGATTTTAGGAGATGGAAGCCAGTTTGAACAGTCGATTCTCTATCATGTAGAGTT 1529
QY 297 -----AlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu 314
Db 1528 TATAAGCCTTCTGCTGATTTGTCTPC-----TTG 1499
QY 315 GluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle 334
Db 1498 CTTCCAGACTTCAAGATAGTTACCAAGAGTTCTCGAAGATGCGACCTATATTCAA 1439
QY 335 SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLys 354
Db 1438 ATGATGACAGGCTTAGATGAGCGACTTTGGCTTTTGGTGATAGCGATTCTACAGAAAG 1379
QY 355 AsnPheArgMetAlaGlnPheAla----- 362
Db 1378 ACAGAAATTTGAGCCTGCTCTGCTGGTTTGAACCCAGAGACACTTCTTAACGGTCTG 1319
QY 363 -----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPhe 377
Db 1318 GATGTTAAAGTTGATTGTGCTTAGCCTCTTGTCTTGGGGCGAGAAAAGTCAAGCGACTG 1259
QY 378 AlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAla 397
Db 1258 CAGGAATTTGAAAAGAGAGCTTGGCGACCTAAG-----TCCATGATCTTTTGAAGACTCT 1205
QY 398 GlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAla 417
Db 1204 GGACATGCTGCATTAAGGAT-----GAACATCGTTATCTATTCTTCAAAAT 1157
```

```
QY 418 SerProProGlyCluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLys 437
Db 1156 GGTCCGCTAGGAGTCCCATAGTCATAGTACGAGAAATAGTTTTGCTTACAGATCA 1097
QY 438 GlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMet 457
Db 1096 GGCCAACTATTCTTATAGATGCTGGCGTTATCTTATCG--GGAGATATATGAACGTTA 1038
QY 458 LysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGln 477
Db 1037 TCTCTAAAGAGTGTGGAGTCATTCGACCTGCATGT----- 999
QY 478 AsnMetValIleThrLysAlaArgGlnAsnLysTrpGlu---ThrGlyAsnAsnLeuAsp 496
Db 998 -----AGATGGAAAGCTCCGAAAGAAATCACGGG 969
QY 497 ValLeu-ThrTyr----- 500
Db 968 ATCCTGGGAATATGAATACTATCTCTCCTCCTGTTTGTCCACATAAAGAGGAGGG 909
QY 501 -----ThrAsnProSerTyrProAsnLeuAspHi 510
Db 908 ATGCAATTATATGAGGGGCTTATGTGCAGAGAACCTGATTTGCCCTTATCTT---CA 852
QY 510 sGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAsp-----ArgAl 528
Db 851 CAGGAGAAAATCCTCATGTTGGTAGAGGATCTCGGCTCTTGGTAGATCATCAGGTG 792
QY 528 alleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnPr 548
Db 791 TCAAGTCAA-----CATGAGGTGTGACTCAGTTT----- 761
QY 548 oValPheAspLysThrLysAsnArgValTyrThrThrTyrArgAspGly-----AsnAs 566
Db 760 -ATCCTTGACAGGAT-----GTGACCTATCAAGATGGGAAAATCAATCA 717
QY 566 nLeuMetIleGlnSer 571
Db 716 GTTGAGACTATGGAGT 701
RESULT 5
US-09-134-001C-795
; Sequence 795, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 795
; LENGTH: 2097
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-795
Alignment Scores:
Pred. No.: 2,66e-05 Length: 2097
Score: 132.50 Matches: 140
Percent Similarity: 34.59% Conservative: 108
Best Local Similarity: 19.53% Mismatches: 260
Query Match: 3.79% Indels: 210
DB: 4 Gaps: 30
US-09-802-285A-2 (1-659) x US-09-134-001C-795 (1-2097)
```

QY 9 IleIleValPheAlaValIleAlaLeuSerSerGlyAsnIle---LeuAlaGlnSerSer 27

Patent No. 5286486
 GENERAL INFORMATION:
 APPLICANT: Payne, Jewel M.
 APPLICANT: Fu, Jenny M.
 TITLE OF INVENTION: No. 5286486el Bacillus thuringiensis Gene
 TITLE OF INVENTION: Encoding a Coleopteran-Active Toxin
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: David R. Saliwanchik
 STREET: 2421 N.W. 41st Street, Suite A-1
 CITY: Gainesville
 STATE: FL
 COUNTRY: USA
 ZIP: 32606
 COMPUTER READABLE FORM: disk
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/973,320
 FILING DATE: 19921106
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/788,638
 FILING DATE: 6-NOV-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Saliwanchik, David R.
 REGISTRATION NUMBER: 31,794
 REFERENCE/DOCKET NUMBER: MA69.C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 904-375-8100
 TELEFAX: 904-372-5800
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3414 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Bacillus thuringiensis
 STRAIN: dakota
 INDIVIDUAL ISOLATE: HD511
 IMMEDIATE SOURCE:
 LIBRARY: Lamdagem (TM)-11 library of J.M. Fu
 CLONE: 511
 US-07-973-320-1

Alignment Scores:
 Pred. No.: 0.000669 Length: 3414
 Score: 123.50 Matches: 141
 Percent Similarity: 35.92% Conservative: 95
 Best Local Similarity: 21.46% Mismatches: 199
 Query Match: 3.53% Indels: 224
 DB: 1 Gaps: 39

US-09-802-285A-2 (1-659) x US-07-973-320-1 (1-3414)
 QY 48 AsnLysAlaValAla-----AlaGlyAsnTyrAspAspAlaAlaLysAlaLeu 63
 Db 361 AATAAGCACTTCGACAAATTAGAGGATTAGGAATAAC----- 399
 QY 64 LeuAlaTyrArgGluLysSerLysAlaArgGluProAspPheSerAsnAlaGluLys 83
 Db 400 TTAACAATATATCAACAG-----GCATTGAAGATTGGCTGAACAATCTGTATGAT 450
 QY 84 ProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAspLysAlaLeu 103
 Db 451 CGAGCACTATACACAGGTAGATCGTTT-----CGTATATAGAT-----GCTTTA 501

QY 104 ValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAsp----- 117
 Db 502 TTGGAATCATATATGCCGTCATTTAGGTTGCTGGATATGAATACCATTTACTAACAGTT 561
 QY 118 TyrGlyLysAspIleAsnTrpGlnMet----- 126
 Db 562 TAGGCACAAGCGGCAACCTTCATCTAGCTTTATTAAAGAGATTCTACTCTTTATGGAGAT 621
 QY 127 ---TrpProValLysAspAsnGluVal----- 134
 Db 622 AAATGGGATTTCATCTAGAGCAACATTTAGGAAAAATTAATTCGTCAAAAGAAACATATC 681
 QY 135 ---ArgTrpGlnLeuHisArgValLysTrpTyrGlnAla---MetAlaLeuValTyrHis 152
 Db 682 TCTGAATATTTAAACCATTTGCGTTAGTGTATATAGTGTCTTAGAGATTGAACGGT 741
 QY 153 AlaThrGlyAspGlu-----LysTyrAlaArgGluTrpValTyrGlnTyr 167
 Db 742 TCACCTTATGAACAATGGATAAATTATAATCGTTTCGTAGAGAAATGATATTAAATGGTA 801
 QY 168 SerAspTrpAlaArgLysAsnProLeu-----GlyLeu 178
 Db 802 TTAGATATTGCTGCTGTATTCTCTATTATGACCTCGAATGTATTCAATGGAAAAAAGT 861
 QY 179 SerGlnAspAsnAspLysPheValTrpArg-ProLeuGluValSerAspArgValGlnSe 198
 Db 862 AGCAGTT-AAAGAGAGAGTGTATACCATTCATTTAGCTTGTCAATTTAGCAATCCAGA 920
 QY 198 rLeuProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMe 218
 Db 921 TATAGTCCAAAGTTTCTTCAGATGGAAAAATCTGCGTTTAGAACACACACCTTGT-- 978
 QY 218 tGluPheLeuAsnSerTyrHisGlnAlaAspTyrLeuSerThrHisTyrAlaGlu 238
 Db 979 -----GATTATTAGTAGCTTTATATATATATAC 1007
 QY 238 nGlyAsnHisArgLeuPhe-----GluAlaGlnArgAsnLeuPhe-----Al 252
 Db 1008 ATCAAAATATAAAGCATTTTCACATGAGATTCAACACGACCTATTATTATGGTGTACA 1067
 QY 252 aGlyValSerPheProGluPheLysAspSerProArgTrpArgGlnThrGlyIle-Serv 272
 Db 1068 TAAGGTAGCTTTAAAAAATCGAGCAATCCCAATTTATAT--ACAACAGGCATATATGG 1124
 QY 272 aLeuAsnThrGluIle-----LysLysGlnValTyrAlaAspGlyMet---GlnPheG 289
 Db 1125 TAAACAAGTGGATATATTTCATCAGAGCATATTCATTAGAGGAAATGATATCTATAG 1184
 QY 289 luLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySera 309
 Db 1185 AACATTAGCAGCTCCATCAGTTGTAGTTTATCCGTATA--CTCAGAATTATGCTGTGCA 1241
 QY 309 laLys----- 310
 Db 1242 GCAAGTTGAGTTTACGGTTAAAGGGCATGTACATATTAGAGGAGATAACAAATATGA 1301
 QY 311 ---ArgValAsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnM 329
 Db 1302 TCTGAGTATGATTCTATTGATCAATTACCCAGACGAGAGAACCAATACACGAAAAATA 1361
 QY 329 etIleMetAlaLeu-----IleSerIleSerLeuProAspTyrAsnThr- 343
 Db 1362 CACTCATCGA-ITATGTCATGTCACAGCTATATCTAAATCAATCCGATTATGATAATG 1420
 QY 344 -----ProMetPheGlyAspSerTrpIleThrAspLysAsnPheAspMetAlaGlnP 361
 Db 1421 CTACTATCCGATCTTT-----TCTTGG-----ACGCATAGAGTGGCGAGT 1462
 QY 361 heAlaSerTrpAlaArgValPheProAlaAsn-----GlnAlaIleLysTyrP 377
 Db 1463 AT-----TACATAGATCTATCCAAACAAATCAAAAAAATTCACAGCTGTAAAAATGT 1516
 QY 377 heAlaThrAspGlyLysGlnGly-----LysAlaProAsnPheLeuSerLysAlaLeuS 395

```
Db 1517 ATAACTAGATGATCTATCTACAGTTGTCAAGGGCTGGATTACAGCTGGAGATTAG 1576
QY 395 eAsnAlaGlyPheTyrThrPheArgSerGlyTrp-----AspLysAsnAlaThrValM 413
Db 1577 TTAAGAGAGGG-----AGTAATGGTTATATAGGATATAAAAGGCTACCGTA- 1623
QY 413 etValLeuLysAlaSerProGlyGlu-----PheHisAlaGlnP 427
Db 1624 -----AACTCACCACCTTCTCAAAATATCGTGTAGAGTTCAGATCGCCACTA 1672
QY 427 roAspAsnGlyThrPheGluLeuPheLeileLysGlyArgAsnPheThrProAspAlaGlyV 447
Db 1673 GTGTTTCTGGACTATTCAACGCTTTATT----- 1701
QY 447 alPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTyrPyrArgGlnThrA 467
Db 1701 ----- 1701
QY 467 rgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnA 487
Db 1702 -----AATGATGAATAGCGCTTCAAAAAAATTTCAAA 1735
QY 487 snLysTrpGluThr-----GlyAsnAsnLeuAspValLeuThrTyr----- 500
Db 1736 GTACTCTAGAACAAATAGGTGAGGAAAGAT---TTAACCTATGTTTCATTTGGATATA 1792
QY 501 -----ThrAsnProSerTyrProAsn-----LeuA 509
Db 1793 TAGAATATTCTACGACCAITCAATTTCCGAATGAGCATCCAAAAATCACTTTTCATTAA 1852
QY 509 spHis-----GlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAspA 527
Db 1853 ACCATTGTGATAACAATTCACCAATTTATAGATTCATCAATCGAATTTATCCCTGTAGAT- 1911
QY 527 rgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerA 547
Db 1912 -----GTAAATTATGATGAAGAAAAA----- 1935
QY 547 snProValPheAspLysThrLysAsnArgValTyrThrTyrThrTyrArgAspGlyAsnL 567
Db 1936 -----CTAGAAAAAGCACAGAAAGCCGTGAATACCTTGTTTACAGAGGAGAAATG 1987
QY 567 euMet-----IleGlnSerLeuAsnAlaAspArgThrSerLeu 579
Db 1988 CACTCCAAAAATAGTGACAGATTATAAGTGACCAGGTTTCAT 2034
```

RESULT 7

```
US-08-328-962-1
; Sequence 1, Application US/08328962
; Patent No. 5536637
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; TITLE OF INVENTION: A NOVEL METHOD FOR ISOLATING CYTOKINES
; TITLE OF INVENTION: AND OTHER SECRETED PROTEINS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Legal Affairs, Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,962
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/045,267
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: McDaniels, Patricia A.
; REGISTRATION NUMBER: 33,194
; REFERENCE/DOCKET NUMBER: GI 5200
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-876-1170
; TELEFAX: 617-876-5851
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; IMMEDIATE SOURCE:
; CLONE: SUC2
; US-08-328-962-1
Alignment Scores:
Pred. No.: 0.000208 Length: 1542
Score: 123.00 Matches: 109
Percent Similarity: 32.01% Conservative: 76
Best Local Similarity: 18.86% Mismatches: 179
Query Match: 3.52% Indels: 214
DB: 1 Gaps: 31
US-08-802-285A-2 (1-659) x US-08-328-962-1 (1-1542)
QY 94 valThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyr 113
Db 4 ATGACAAACGAACTAGCAGTACCTTTGGTCCAC---TTACACCCCAACAAAGGGCTGG 60
QY 114 -----GlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrp--- 127
Db 61 ATGAATGACCCAAATGGTGTGGTACGATGAAAGAGATGCCAAATGGCATCTGTACTTT 120
QY 128 ProValLysAspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMet 147
Db 121 CAATACACCCAAATGACACCGTATGG-----GGTAGCCATTTGTTGGGCG--- 168
QY 148 AlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpValTyrGlnTyr 167
Db 169 -----CATGTACTCCGATGAT-----TTG 189
QY 168 SerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTrp 187
Db 190 ACTAATGG---GAAGATCAACCCATTGCTATCGCTCCCAAGCGTAAC----- 234
QY 188 ArgProLeuGluValSerArgValGlnSerLeuProProThrPheSerLeuPheVal 207
Db 234 ----- 234
QY 208 AsnSerProAlaPheThrProAlaPheLeuMetGluPhe----- 220
Db 235 GATTCAAGGTCTTCTCTCGCTCCATGGTGGTGGTATTACAAACACGACGATGGGTGTTTC 294
QY 221 -----LeuAsnSerTyrHis 225
Db 295 AATGATACTATTGATCCAGACAAAGATGGTGGATTTGGACTTATACACTCTCTGAA 354
QY 226 GlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGlu 245
Db 355 AGTGAAGAGCAATACATTAGC-----TATTCTCTTGTAGTGTGTACACTTTTACTGAA 408
QY 246 AlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAsp----- 261
Db 409 TACCAAAAGAACCCCTGTTTAGCTGCCAATCCCACTCAATTCAGATCCAAAGGTGTTTC 468
QY 262 -----SerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIle 277
```

```
Db 469 TGCTATGAACCTTCTCAAAATGGAATATGACG-----GCTCCCAAAATCACAAGACTAC 522
QY 278 LysLysGlnValTyrAlaAsp----- 284
Db 523 AAAATTGAAATTTACTCTCTCTGTAAGTCTGGAAGCTAGAATCTGCAATTGCTC 582
QY 285 -----GlyMetGlnPheGluLeu----- 290
Db 583 AACGAAGTTTCTTAGCTACCAATACGAATGTCCAGGTTTATTGAAGTCCCACTGAG 642
QY 291 -----SerProIleTyrHisValAlaAlaAspIle----- 301
Db 643 CAAGATCCTTCCAAATCTTATTGGTGCATGTTATTCTATCAACCCAGGTGCACCTGCT 702
QY 302 -----PheLeuLysAlaTyr 306
Db 703 GCGGTTCTTCAACCAATATTGTTGGTGCATCTTCAATGGTACTCATTTTGAAGCGTTT 762
QY 307 GlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyr-----ValGln 324
Db 763 GACAATCAATCTAGAGTG-----GTAGATTGTTGTAAGACTACTATGCTTGCA 813
QY 325 ThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnThrPro 344
Db 814 ACTTTCTTCAAC-----ACTGACCAACCTACGTTTCAGCA 849
QY 345 MetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPheAla----- 362
Db 850 TTA-----GGTATTGCTTGGCTTCA-----AACTGGGAGTACAGTGCCTTTGCCAACT 900
QY 363 -----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLys 375
Db 901 AACCCATGAGATCATCGATGCTTTGGTCCGCAAGTTTCTTTGNAC-----ACTGAA 954
QY 376 TyrPheAla-----ThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLys 392
Db 955 TATCAAGCTAATCCAGAGACTGAATTGATCAATTTGAAAGCCGACCAATATTGAAC--- 1011
QY 393 AlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrVal 412
Db 1012 ---ATTAGTAATGCTGGT-----CCCTGGTCTGCTTTTCTACTAAC 1050
QY 413 MetValLeuLysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPhe 432
Db 1051 ACAACTCTAACTAAGGCCAAATCTTACAAATGCGATTTGAGCAACTCGACTGCTACCTA 1110
QY 433 GluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGly 452
Db 1111 GAGTTTGAGTGTG-----GTTTACGCTGTTAACACC 1140
QY 453 AspGluAlaIleMetLys-----LeuArgAsnTrpTyrArgGlnThrArg 467
Db 1141 ACACAAACCATATCCAAATCCGCTCTTTCGCGACTATCATCTTGGTTCGAAGGTTTAGA 1200
QY 468 -----IleHisSerThrLeuThrLeuAsp 475
Db 1201 GATCTCTGAAGAATATTTGAGAATGGGTTTGAAGTCAGTGCCTCTTCTCTTTTGGAC 1260
QY 476 AsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGly----- 492
Db 1261 CGTGTAATCTTAAGGTCAAGTTTGTCAAGGAGAACCCATATTTCCACAAACAGAAATGCT 1320
QY 493 ---AsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrProAsn----- 507
Db 1321 GTCACAAACCAACCATTCAGTCTCGAGACCACTAAGTTACTATATAAGTGTACGGCCTA 1380
QY 508 LeuAspHisGlnArgSerValLeuPheIleAsn-----LysLys 520
Db 1381 CTGGATCAAAACATCTTGGAATGTACTTCAACGATGGAGATGTGGTTTCTTACAAATACC 1440
QY 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyVal 538
Db ----- 538
```

```
Db 1441 TACTTTCATGACCACCGGTAACGCTCTAGGATCTGTGACATGACCACCTGGTGTC 1494
RESULT 8
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
Alignment Scores:
Pred. No.: 185 Length: 4403765
Score: 120.50 Matches: 85
Percent Similarity: 34.2% Conservative: 60
Best Local Similarity: 20.0% Mismatches: 150
Query Match: 3.45% Indels: 129
DB: 4 Gaps: 22
US-09-802-285A-2 (1-659) x US-09-103-840A-2 (1-4403765)
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAla----- 148
Db 3679590 GACAAGGCTATCGAGTGG---CTGCACACCGGTGCGCGGCCAGATGCCACCAAGCGGTGG 3679534
QY 149 LeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpValTyrGlnTyrSer 168
Db 3679533 ATGCTGTACTACCGCACCGGC---GCCACCCACGCGCCACACCATCATTTCAAG----- 3679483
QY 169 AspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTyrArg 188
Db 3679482 GAATGGCGCGCAACAGTACCGAGGTGAGTTTCGATGATGCTGGATGTGTACCGCGCAGAAG 3679423
QY 189 ProLeuGluValSerAspArgValGlnSerLeuProProThrPheSerLeuPheValAsn 208
Db 3679422 ACATTCGAACCGCAAAACGACTCGGATCATTCACCCGACGCGCAACTC---ACCGAG 3679366
QY 209 SerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnGlnAla 228
Db 3679365 CGGCCCGACCTATTTCCCGCGGTGG----- 3679342
QY 229 AspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGluAlaGlnArg 248
Db 3679341 GACAGTATGTCG-----GAGGCGCAAAACCGCTCTTTGCCCGCCAGATG 3679297
QY 249 AsnLeuPheAlaGlyValSer----- 255
Db 3679296 GAGGTGTTCCGCGGTTCCTCGAAATGCGGACTGGAATGTTGGCCGCTCTGACACGCG 3679237
QY 256 PheProGluPheLysAspSerProArg-----TrpArgGlnThrGly 269
Db 3679236 ATCCGAGGATCTCGGCGAGTCCGACACACAGTGTGTTCTTACATCTGGGGCCACAAATGGC 3679177
QY 270 IleSerValLeuAsnThr-----GluIleLysLysGlnValTyrAlaAspGlyMet 286
Db 3679176 GCCAGCATGGAGGCGACCAACACCGGTTTCGTTCAATGATGATGACGCTTCTTAAACGCGCTG 3679117
```

```
QY 287 -----GlnPheGluLeuSerProIleTyr---HisValAlaAlaIle 299
Db 3679116 GATCTGGATGCCGAGCGGATTTGGAGCTGATCAACAATATTCGGCGCATCGCGCACTC 3679057
QY 300 -----AspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsn 313
Db 3679056 GCGGACGAGTTACCGCACCAGCATTTGCCAGCGCGTGGCGCATCGGACACACCCCG 3678997
QY 314 LeuGlu-----LysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMet 331
Db 3678996 TTGCAGTGGGCAAGCAGATGGCCACCCTGGCGCGCAGCGCGATCCATTCATTTGGTGGTC 3678937
QY 332 Ala-----LeuIleSerIleSerLeuProAsp----- 332
Db 3678936 GCTTGGCGCGCGCGGATCCGCGCAGACGCGCGTTCGTAGCCAGTTCCACCCTGCATC 3678877
QY 333 -----LeuIleSerIleSerLeuProAsp----- 340
Db 3678876 GACATCGCGCGCGCGGATTTGGCGGCATCGTTTACCGAGCGCGACCCATGTCGACGGC 3678817
QY 341 ---TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAla 359
Db 3678816 TTCGACGAGCAACCGATGGACGGAACCACTTCGTG----- 3678781
QY 360 GlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAla-----Ile 374
Db 3678780 -----CGGACCTTCGACGAGCTGAAGCCGGAAGACCGCCACACCGCTG 3678739
QY 375 LysTyrPheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeu 394
Db 3678738 CAGTACTTC-----GAAACTTCGCGACCGTGCATC 3678706
QY 395 SerAsnAlaGlyPheTyrThrPhe-----ArgSerGlyTrpAspLysAsnAla 410
Db 3678705 TACAAAGACGCGTGGTGGCGTGGCTGGTTCGACAGCGCGCTGGGATCTGTCAACG 3678646
QY 411 ThrValMetValLeuLysAlaSerProProGlyGluPheHisAlaGlnProAspAsnGly 430
Db 3678645 GAGACAGATG-----CGACGGTTTCGCGCGCGGACCTAC-----GACCGGACCGAGAC 3678598
QY 431 ThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyr 450
Db 3678597 GTCCTGGAGCTGTACTACCTACAGATGACTTCTCCAGCGGAAAAACCTGGCAGCGGAG 3678538
QY 451 SerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSer 470
Db 3678537 CATCCCGACAAGTCCGCGAGCTCACCAGCTGTGGTGGCAGGAGCGGACGAAACCGG 3678478
QY 471 ThrLeuThrLeu 474
Db 3678477 GTGCTCGCGCTG 3678466

RESULT 9
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
```

OTHER INFORMATION: H37Rv

US-09-103-840A-1

Alignment Scores:

Pred. No.: 186 Length: 4411529

Score: 120.50 Matches: 85

Percent Similarity: 34.20% Conservative: 60

Best Local Similarity: 20.05% Mismatches: 150

Query Match: 3.45% Indels: 129

DB: 4 Gaps: 22

US-09-802-285A-2 (1-659) x US-09-103-840A-1 (1-4411529)

QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpGlnAlaMetAla----- 148

Db 3684804 GACAAAGCTATCGAGTGG---CTGCACACCGTGGCGGCCGACCAATCCCAAGCGGTGG 3684748

QY 149 LeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpValTyrGlnTyrSer 168

Db 3684747 ATGCTGTACTACGCGACCGGC---GCCACCCAGCGCCACACACCGTATTCAG----- 3684697

QY 169 AspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTrpArg 188

Db 3684696 GAATGGCGCGCAAGTACCGAGTGGTTCGATGATGGTGGGATGTGTACCGCGAGAAG 3684637

QY 189 ProLeuGluValSerAspArgValGlnSerLeuProProThrPheSerLeuPheValAsn 208

Db 3684636 ACATTCGAAGCGCAAAAGCGACTCGGGATCATTCACACCGCGACCGCAACTC---ACCGAG 3684580

QY 209 SerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnAla 228

Db 3684579 CGGCCGACCTATTCGCCGCTGG----- 3684556

QY 229 AspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGluAlaGlnArg 248

Db 3684555 GACAGTATCTCG-----GAGGCGCAAAACGGCTCTTTGCCGCCAGATG 3684511

QY 249 AsnLeuPheAlaGlyValSer----- 255

Db 3684510 GAGTGTTCGCGCGGTTCCTCGGAAATCGGACATGGAAATGTTGGCGCGTGTGGACGG 3684451

QY 256 PheProGluPheLysAspSerProArg-----TyrArgGlnThrGly 269

Db 3684450 ATCGAGGATCTCGGCGAGTCCGACACACAGTGGTGTCTTACATCTGGCGCGCATGGC 3684391

QY 270 IleSerValLeuAsnThr-----GluIleLysLysGlnValTyrAlaAspGlyMet 286

Db 3684390 GCCAGCATGGAGGGCACCAACACCGGTTCTCAATGAGATGACGTTCTTAAACGGCTG 3684331

QY 287 -----GlnPheGluLeuSerProIleTyr---HisValAlaAlaIle 299

Db 3684330 GATCTGGATGCCGAGCGCAATTTGGAGCTGATCGAACAATACGCGCGCATCGCGCACTC 3684271

QY 300 -----AspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsn 313

Db 3684270 GCGCAGAGTTACCGCACCAGCATTTCCGCGCGGTGGCGCATCGGACACACACCCCG 3684211

QY 314 LeuGlu-----LysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMet 331

Db 3684210 TTGCAGTGGGCAAGCAGATGGCCACCCTGGCGCGCAGCGCGATCCATTCATTTGGTGGTC 3684151

QY 332 Ala----- 332

Db 3684150 GCTTGGCGCGCGGATCCGCGCAGACGCGCGTTCGTAGCCAGTTCCACCCTGCATC 3684091

QY 333 -----LeuIleSerIleSerLeuProAsp----- 340

Db 3684090 GACATCGCGCGCGCGTGTGGCGGCATCGTTTACCGAGCGCGACCCATGTCGACGGC 3684031

QY 341 ---TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAla 359

Db 3684030 TTCGACGAGCAACCGATGGACGGAACCACTTCGTG----- 3683995

	Qy	104	ValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAsp-----	117
	Dq	502	TTTGAATCATATATGCCCGTCATTAGGGTTCTCGGTATGTAATAACCACTTACTAACAGTT	561
	Qy	118	TyrGlyLysAspIleAsnTrpGlnMetTrpProValLysAspAsnGluVal-----	134
	Dq	562	TACGCACAAGCGGCAAACTTCATCTAGCTTTATTAAAGAGATTCTACTCTTTTAGGAGAT	621
	Qy	135	ArgTrpGlnLeu-----	138
	Dq	622	AAATGGGAATTCCTCAGAACAACATTGAGGAAAAATTATAATCGTCAAAAAGAAACATATT	681
	Qy	139	-----HisArgValLysTrpTrpGlnAla---MetAlaLeuValTyrHis	152
	Dq	682	TCTGAATATTCTAACCATTCCGTTAAGTGATTAATAGTGTCTTTAGCAGATTGAACGGT	741
	Qy	153	AlaThrGlyAspGlu-----LysTyrAlaArgGluTrpValTyrGlnTyr	167
	Dq	742	TCCACTTATGAACAACATGGATAAAATTATAATCGTTTTCTGAGAGAAATGATTAATGGTA	801
	Qy	168	SerAspTrpAlaArgLysAsnProLeu-----GlyLeu	178
	Dq	802	TTAGATATTCTGCTGTATTTCCTTATTATGACCCCTCGAAATGTATTCATATGAAACAAGT	861
	Qy	179	SerGlnAspAsnAspLysPheValTrpArg-ProLeuGluValSerAspArgValGlnSe	198
	Dq	862	ACGCAGTT-AACGAGAGAAGTGATATACCGATCCAATTAGCTTGTCAATTAGCAATCCAGG	920
	Qy	198	rLeuProProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMe	218
	Dq	921	TATAGTCCCAAGTTTTTCTCAGATGGAAAACTACGCGATTAGAACACCCACACCTTGT--	978
	Qy	218	tGluPheLeuAsnSerTyrrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGl	238
	Dq	979	-----GATTATTTAGATGAGCTTTTATATATATAC	100
	Qy	238	nGlyAsnHisArgLeuPhe-----GluAlaGlnArgAsnLeuPhe-----Al	252
	Dq	1008	ATCAAAATATAAGCATTTTTCATGATGAGATTCACCCAGACCTATTTTATTCGGATGCACA	106
	Qy	252	aGlyValSerPheProGluPheLysAspSerProArgTrpArgGlnThrGlyIleSerVa	272


```
QY 292 ProIleTyrHisValAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArg 311
Db 460 CATTTCCAAACAGATGACAAATTCGTATTT-----TGTCAGAGAGAACACACGGA 510
QY 312 ValAsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIle--- 330
Db 511 CTTATGCTAGAT---CAACACACACGCTTGTTACTCTACTACACACCACTTAATTTCT 567
QY 331 -----MetAlaLeuIleSerIleSerLeuProAspTyrAsn 342
Db 568 GTAAGAGAAATCTTTAAAGATGTCGAACTACAAATTGAAACAGCTTTACACATTTTAA 627
QY 343 ThrProMetPheGlyAspSerIleThrAspLysAsnPheArgMetAlaGlnPheAla 362
Db 628 AATCTTATCTATCC----- 642
QY 363 SerTTPAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAspGlyLys 382
Db 643 -----ACAAATGATTAACACTCCAGATTTTAAATAGCGACACTGTC 681
QY 383 GlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsn-----AlaGly 398
Db 682 TTATTAACTTACCAATGATATACACAGCGAGCTTTAAATCTTATATTACATCAAGG 741
QY 399 PheTyrThrPheArgSerGlyTTPAspLys-----AsnAla 410
Db 742 TATATTCAATTCGCAAAAGATGGAATCTGTAATTTATGATGAAAGCTTTTATAATCAA 801
QY 411 ThrValMetValLeuLysLaserProProGlyGluPheHisAlaGlnProAspAsnGly 430
Db 802 ACAAAGATTGATTTGCAAGCTGCTATT-----CAGGACTATTCTACTACTGTATCTACC 855
QY 431 ThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyr 450
Db 856 ACTTTTGAATAATTCAAACCTACTCTAAAT-----CCA 888
QY 451 SerGlyAspGluAlaIleMetLysLeuArgAsnTyrTyrArgGlnThrArgIleHisSer 470
Db 889 TCAATAAAGAATCTGTTAATAGTATATAGATATGTCGTTCCTGACTCTTCAATCT 948
QY 471 Thr-----LeuThrLeuAspAsn----- 476
Db 949 TTAGACATTCGTCACATCGGCTACTTTAGATAATGTTAATTAACCTTCCAAATGATAG 1008
QY 477 -----GlnAsnMetValIleThrLysAlaArgGlnAsnLysTyrGlu----- 490
Db 1009 ATTCAATTGGATCAAACTCGCTTAGTATTTTCAGATGTTGCAGGACCTTGGGAAGTAA 1068
QY 491 -----ThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnPro----- 503
Db 1069 GATAATATAACTTCGAATATTAGATGATTATTAACACCAATAAATACAGGATAGATT 1128
QY 504 -----SerTyrProAsnLeuAspHisGln----- 511
Db 1129 CAAGAAAGTTCAGATCTTAGAAATTTCACTTATCCAGTAAGATTACAAAGCATGCAA 1188
QY 512 -----ArgSerVal-----LeuPhe 516
Db 1189 TTCCATGGCAATATGTAACCTCAAAAGTGTAGAACATTTATAGCGATGCTCTTAA 1248
QY 517 IleAsnLysLysTyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeu 536
Db 1249 TTAATAATATAA-----AATAAACTATAACTGCAGGTGTAAGTAAATAT 1293
QY 537 GlyValHisTyrGlnLeuLysGluAspSerAsn----- 547
Db 1294 -----GATGAAGTAACTCAAAATAATAAACAATAACTATGCTCT 1332
QY 548 -----ProValPheAspLysThrLysAsnArgValTyrThrThrTyrArgAsp 563
Db 1333 GTAATAAATAGCTTACTGATATCAAGTAAATTCACAAATTTCTCAATATTAGAT 1392
```

```
QY 564 GlyAsnAsnLeuMetIleGln-----SerLeu 572
Db 1393 TTAATTTCTAGTCATGTAATAATGGTGTCAAAAAGTAGCGGGTGTTCACCACTTAGTTCA 1452
QY 573 AsnAlaAspArgThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsnLys 592
Db 1453 AATGGTAATTTCTAATAATGCTGTTTACCTAATCAAAATAAATGTTATTATTTCAGTA 1512
QY 593 GluLeuLysArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGln--- 611
Db 1513 CAATCAAAAT-----GATAAACCCAGAAAAAATCATGCAGACACTTATAGA 1554
QY 612 -----AsnPheValSerIleValTyrProTyrAsp----- 621
Db 1555 AATGGGATATATAGACAGCTCATATTCCTTTATGATCTTGTCCAGAAAATGTAATGGA 1614
QY 622 -----GlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAsp 637
Db 1615 GATATAGATCCGATACTAAACACCGTCATCTGCTCTT-----AAAGGTTTCCG 1665
QY 638 PheGluLysGly 641
Db 1666 GCAGAAAAAGGA 1677
RESULT 12
PCT-US96-05320A-637
; Sequence 637, Application PC/TUS9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams
; APPLICANT: Owen White
; APPLICANT: Hamilton O. Smith
; APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20003-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05320A
; FILING DATE: April 22, 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: June 7, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Eric K. Steffe
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.014PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 637:
; SEQUENCE CHARACTERISTICS:
```



```

Db      2140 ATTGATAAAGTGCT-----GCATAGATATACCGTTTTATCAAAACCAAAAT 2187
Qy      624 -----LysAlaProGluIleSerIleArgGluAsnLysGlyAsnAsp 637
Db      2188 CGAGATCAAGCAGAGAGTCAGAGGTATCGAAATTGCTTCAGCTCTTGAATGGGTGATTG 2247
Qy      638 PheGluLys 640
Db      2248 TTTGAAAAA 2256

RESULT 13
US-08-961-527-74
; Sequence 74, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-74

Alignment Scores:
Pred. No.: 0.0874 Length: 16535
Score: 115.00 Matches: 161
Percent Similarity: 33.03% Conservative: 94
Best Local Similarity: 20.85% Mismatches: 263
Query Match: 3.29% Indels: 254
DB: 4 Gaps: 45

US-09-802-285A-2 (1-659) x US-08-961-527-74 (1-16535)

Qy      11 ValPheAlaValIleAlaLeuSerSerGly-----AsnIleLeuAla 24
Db      1387 GTAAGTGGAACTCGCTCTTGAAAAAGGGGAAAGAGGAACCTTGGAAACAAACTCTAGAC 1446
Qy      25 GlnSerSerIleThrArgLysAspPheAspHisIleAsnLeuGluTyrSerGlyLeu 44
Db      1447 AGCACAACAACTCGGAATCACAGATTTCATCGGCTACTATATCAATACCA---ATC 1503
Qy      45 GlnLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAlaAlaLysAlaLeu 64
Db      1504 GAGCGTCAAGGTAAACGTGTTCTTGA-----CTCGATCCCTACGCTAAATCTCTGCT 1557

```

```
QY 369 ProAlaAsnGlnAlaIleLysTyrPheAlaThrAspGlyLysGlnGlyLysAlaProAsn 388
Db 2416 ---CATACGATCTGCTGCTGCTTTTTCAGATGACATCCG---AAC 2457
QY 389 PheLeuSerLysAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTyrAspLys 408
Db 2458 AACCTCAAAATCTGTTATCCAAACGAAGGTCAACCTGCCTTTATCACAGGTGGCAAGCGT 2517
QY 409 AsnAlaThrValMetVal-----LeuLysAlaSerPro 419
Db 2518 GATGTCAACACCATCTTTAAATCTCATGCTCAACCACTAACTTTGAAGCTGACAGC 2577
QY 420 ProGly-----GluPheHisAlaGlnProAspAsnGlyThr---PheGluLeuPhe 435
Db 2578 CCTGGAGATGTCATCCATATACCGCAGCCCATGTAACCTTGACCTCTTTTGACATCAT 2637
QY 436 Ile-----LysGlyArgAsnPhe-----441
Db 2638 GCCCAGTCTCAAAAAGACCAAGCAAGGCTGAGAACTATGCTGAAATCCACCGTCTG 2697
QY 442 -----ThrProAspAlaGlyValPheValTyrSerGly 452
Db 2698 TTACGACTTGGAAATCTCATGCTTTCAGACCTCAAGAACTCCATTTATCCACTCCGGT 2757
QY 453 AspGluAlaIleMetLysLeuArgAsnTyrTyr-----ArgGlnThrArgIleHis 469
Db 2758 CAGGAA-----TATGGAGTCTATAACAATCCGTGACCCA 2793
QY 470 SerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrp 489
Db 2794 GCCTCAAGACTCCAGTAGCAGAGATAAGGTTCCAAACAAATCTCACTTGTTCGGTGAT 2853
QY 490 GluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrProAsnLeuAsp 509
Db 2854 AAGGCGGCAACCACTTGTGATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2913
QY 510 HisGlnArgSerValLeuPheIleAsnLysTyrPheLeuValIleAspArgAlaIle 529
Db 2914 -----GCAGTCAACAG-----2925
QY 530 GlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnProVal 549
Db 2926 -----TTTGACTGGACTAAGGCTACAGATGCTAAAGCTTAT 2961
QY 550 PheAspLysThrLysAsnArgValTyrThrTyrArgAspGlyAsnAsnLeuMetIle 569
Db 2962 CCTGAAATGTCAGAGCCGT-----GACTATATGAAAGGTTTGTATGTCCTTCGT 3012
QY 570 GlnSerLeuAsnAla-----AspArgThrSerLeu 579
Db 3013 CAATCTACAGATGCTTCCGACTTAGAGTCTTCAAGATATCAAGACCGGTGCCACCTC 3072
QY 580 AsnGluGluGly-----LysValSerTyrValTyrAsnLysGluLeu 594
Db 3073 ATCACTGTCAGGCCAAATGTTGGAAAAGAGGATGTAGTGTGGTCTACCAATC 3132
QY 595 LysArgPro-----AlaPheVal-----PheGluLysProLysLys 606
Db 3133 ACTGCTCCAAACGCGCATCTACGCGACTCTTTGTCAATGCGGATGAAAAGCTCGGAA 3192
QY 607 ---AsnAlaGlyThr-----GlnAsnPheValSerIleValTyrProTyr 620
Db 3193 TTATTTGGAACTGCTTTCACATCTAAGAAATCGGAAGTTTGGCA-----3243
QY 621 AspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAspPhe---Glu 639
Db 3244 GATGAAAACCAAGCAGCAGCAGTCCGGAATTCGCAACCCGAAAGGACTTGAATGACTGAA 3303
QY 640 LysGly---LysLeuAsn---LeuThrLeuThrIle 649
Db 3304 AAAGCTTGAATTTGAATGCTTTCACGCTACTGTT 3339
```

RESULT 14

```
US-08-485-568A-5
; Sequence 5, Application US/08485568A
; Patent No. 5589382
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Fu, Jenny M.
; TITLE OF INVENTION: No. 5589382el Bacillus thuringiensis Genes Encoding
; TITLE OF INVENTION: Nematode-Active Toxins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,568A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,197
; FILING DATE: 21-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/918,345
; FILING DATE: 21-JUL-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/558,738
; FILING DATE: 27-JUL-1990
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,698
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,403
; FILING DATE: 30-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/999,053
; FILING DATE: 14-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA48DD2.C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3504 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-485-568A-5
Alignment Scores:
Pred. No.: 0.00789
Score: 114.50
Percent Similarity: 31.49%
Best Local Similarity: 19.66%
Query Match: 3.28%
DB: 1
Length: 3504
Matches: 103
Conservative: 62
Mismatches: 170
Indels: 189
Gaps: 24
```

US-09-802-285A-2 (1-659) x US-08-485-568A-5 (1-3504)

249 AsnLeuPheAlaGlyValSerPheProGluPhe-----LysAspSerProArgTyrArg 265
 280 AATATGTTGTAGAGCTTTATTTTCCGAAATATTTCGAGAGAGGTCTCTCAACAAATGCT 339
 267 GlnThrGlyIleSerValLeuAsnThrGluIleLysGlnVal----- 281
 340 CAAGCTCAATTTTTCGAACTAATCATAGAAAAGTTTAAAGAACTTGTGTGATGAAGATTTT 399
 282 -----TyrAlaAspGlyMetGlnPheGluLeuSer 291
 400 AGAAACTTTTACCCTTAATACTACTCAATTTACCTTGTATGCTATGCAACAGCCTTATCA 459
 292 ProIleTyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArg 311
 460 CATTTCCAAAACGAGTACAAATTTGCTATT-----TGTCAAGGAGAACCAACACGGA 510
 312 ValAsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIle--- 330
 511 CTTATGCTAGAT---CAAACACACGCGCTGTACTCTACTACAGACCAATTTAATTTCT 567
 331 -----MetAlaLeuIleSerIleSerLeuProAspTyrAsn 342
 568 GTAAGAGAAATCTTTTAAAGATGCTCGAACTACAAATTTGAAACAGCTTTTACCACATTTTAAA 627
 343 ThrProMetPheGlyAspSerTyrIleThrAspLysAsnPheArgMetAlaGlnPheAla 362
 628 AATCCTATGCTATCC----- 642
 363 SerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAspGlyLys 382
 643 -----ACAAATGATAACACTCCAGATTTTAAATCTTATATACATCAAGGG 741
 383 GlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsn-----AlaGly 398
 682 TTATTAACATTACCAATGTATACACAGGAGCGACTTTTAAATCTTATATACATCAAGGG 741
 399 PheTyrThrPheArgSerGlyTyrAspLys-----AsnAla 410
 742 TATATTCAATTCGAGAAAGATGAAATCTGTAATTTATGATGAAAGTTTATATATCA 801
 411 ThrValMetValLeuLysAlaSerProProGluPheHisAlaGlnProAspAsnGly 430
 802 ACAAAGTTGATTTGCAACGCTGCTATT-----CAGGACTATTCTACTGCTATCTACC 855
 431 ThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyr 450
 856 ACTTTTGAAAAATCAACCTACTCTAAAT-----CCA 888
 451 SerGlyAspGluAlaIleMetLysLeuArgAsnTyrTyrArgGlnThrArgIleHisSer 470
 889 TCAATAAAGAATCTGTATAAAGTAAATAGATATGTCGTTCCATGACTCTTCAATCT 948
 471 Thr-----LeuThrLeuAspAsn----- 476
 949 TTAGACATTGCTGCAACATCGCCTACTCTTAGATAATGTTAAATTACCCCTTCCAATGTAGAT 1008
 477 -----GlnAsnMetValIleThrLysAlaArgGlnAsnLysTyrGlu----- 490
 1009 ATTCAATTTGGATCAAACTCGCTTAGTATTTTCAGATGTTTCAGGACCTTTGGGAAGTAA 1068
 491 -----ThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnPro----- 503
 1069 GATAATAATAATTCGAATATTAGATGTTTAAACACCAATAATAACAGGATAGGATTT 1128
 504 -----SerTyrProAsnLeuAspHisGln----- 511
 1129 CAAGAAAGTTTCAGATCTTAGAAAATTCACCTATCCAGATAGAAATTCACAAACATGCAA 1188
 512 -----ArgSerVal-----LeuPhe 516

1189 TTCCATGGACAATATGTAACCTCAAAAAGTGTAGAACACTGTTTATAGCGATGCTTTAAA 1244

517 IleAsnLysLysTyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeu 536
 1249 TTAATAATTATAAA-----AATAAACTATAACTGCAGGTGTAAAGTAATATT 1293
 537 GlyValHisTrpGlnLeuLysGluAspSerAsn----- 547
 1294 -----GATGAAGTAATCAAAATTAATAAAACATAACTATGCTCT 1332
 548 -----ProValPheAspLysThrLysAsnArgValTyrThrThrTyrArgAsp 563
 1333 GTAATAAATAGTCTTACTGATATCAACGTAATATCCCAAAATCTCAATATTAGAT 1392
 564 GlyAsnAsnLeuMetIleGln-----SerLeu 572
 1393 TTAATTTCACTCATGTATAATCGTGTGTGTCACAAAGTAAACCGGTGTTCCACCACTTATTTCA 1452
 573 AsnAlaAspArgThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsnLys 592
 1453 AATGGTAATCTTAATAATGCTCTTACCTAATCAAAATTAATGTTTATTTATTCAGTA 1512
 593 GluLeuLysArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGln--- 611
 1513 CAATCAAAAT-----GATAAACCAAGAAACATGCAGACACTTATAGA 1554
 612 -----AsnPheValSerIleValTyrProTyrAsp----- 621
 1555 AAATGGGATATATGAGCAGTCAATATCTTATGATCTTTCCAGAAAATGTAATTGGA 1614
 622 -----GlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAsp 637
 1615 GATATAGATCCGAGTAACTAAACACCGTCAATGCTCTCTT-----AAAGGGTTCCG 1665
 638 PheGluLysGly 641
 1666 GCAGAAAAGGA 1677

RESULT 15
 US-08-590-554A-5
 ; Sequence 5, Application US/08590554A
 ; Patent No. 5831011
 ; GENERAL INFORMATION:
 ; APPLICANT: Payne, Jewel
 ; APPLICANT: Narva, Kenneth E.
 ; APPLICANT: Fu, Jenny M.
 ; TITLE OF INVENTION: No. 5831011el Bacillus thuringiensis Genes Encoding
 ; TITLE OF INVENTION: Nematode-Active Toxins
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: David R. Saliwanchik
 ; STREET: 2421 N.W. 41st Street, Suite A-1
 ; CITY: Gainesville
 ; STATE: Florida
 ; COUNTRY: USA
 ; ZIP: 32606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/590.554A
 ; FILING DATE: 21-MAR-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/485,568
 ; FILING DATE: 7-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/310,197
 ; FILING DATE: 21-SEP-1994
 ; CLASSIFICATION: 435

Qy 622 -----GlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAsp 637
Db 1615 GATATAGATCCGGATACCTAAACACCGTCATTGCTTCTT-----AAAGGGTTTCCG 1665
Qy 638 PheGluLysGly 641
Db 1666 GCAGAAAAAGGA 1677

Search completed: August 4, 2003, 14:14:52
Job time : 6299 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame plus p2n model

Run on: August 4, 2003, 07:41:04 ; Search time 405 Seconds
(without alignments) 3664.363 Million cell updates/sec

Title: US-09-802-285A-2
 Perfect score: 3494
 Sequence: 1 MTTKIFKRIIVFAVIALSSG.....KGKLNLTITNGKQQLVLVP 659

Scoring table:	BLOSUM62	
	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

```
Q/cgcn2.1/USPTO_spool/US09802285/runat 28072003 190510 9692/app_query.fasta_1.839
DBEN Genesec 101002 -QMT1-fsrapcat -SUFFIX=ring -MINMATCH=0.1 -LOOFCNT=0
-LOOPEXT=0 -LIST=bits -START=1 -END=1 -MATRIX=blomsumc2 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORES=pct -THR MAX=100 -THR MIN=0 -ALIGN=45
-MODE=LOCAL -OUTFMT=ptho -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=20000000000
-USER=US09802285 @CGN 1 263 @runat 28072003 190510 9692 -NCPU=6 -ICPU=3
-NO MMAP -LARGEXJECT -NEG SCORES=0 -WAIT -DSPBLOCKS=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : N Genes 101002: *

Database :	
1: /SID52/cgdata/geneseq/geneseqn-emb1/NA1980.DAT:*	
2: /SID52/cgdata/geneseq/geneseqn-emb1/NA1981.DAT:*	
3: /SID52/cgdata/geneseq/geneseqn-emb1/NA1982.DAT:*	
4: /SID52/cgdata/geneseq/geneseqn-emb1/NA1983.DAT:*	
5: /SID52/cgdata/geneseq/geneseqn-emb1/NA1984.DAT:*	
6: /SID52/cgdata/geneseq/geneseqn-emb1/NA1985.DAT:*	
7: /SID52/cgdata/geneseq/geneseqn-emb1/NA1986.DAT:*	
8: /SID52/cgdata/geneseq/geneseqn-emb1/NA1987.DAT:*	
9: /SID52/cgdata/geneseq/geneseqn-emb1/NA1988.DAT:*	
10: /SID52/cgdata/geneseq/geneseqn-emb1/NA1989.DAT:*	
11: /SID52/cgdata/geneseq/geneseqn-emb1/NA1990.DAT:*	
12: /SID52/cgdata/geneseq/geneseqn-emb1/NA1991.DAT:*	
13: /SID52/cgdata/geneseq/geneseqn-emb1/NA1992.DAT:*	
14: /SID52/cgdata/geneseq/geneseqn-emb1/NA1993.DAT:*	
15: /SID52/cgdata/geneseq/geneseqn-emb1/NA1994.DAT:*	
16: /SID52/cgdata/geneseq/geneseqn-emb1/NA1995.DAT:*	
17: /SID52/cgdata/geneseq/geneseqn-emb1/NA1996.DAT:*	
18: /SID52/cgdata/geneseq/geneseqn-emb1/NA1997.DAT:*	
19: /SID52/cgdata/geneseq/geneseqn-emb1/NA1998.DAT:*	
20: /SID52/cgdata/geneseq/geneseqn-emb1/NA1999.DAT:*	
21: /SID52/cgdata/geneseq/geneseqn-emb1/NA2000.DAT:*	
22: /SID52/cgdata/geneseq/geneseqn-emb1/NA2001A.DAT	
23: /SID52/cgdata/geneseq/geneseqn-emb1/NA2001B.DAT	
24: /SID52/cgdata/geneseq/geneseqn-emb1/NA2002.DAT:*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	3494	100.0	1980	17	AAQ99227	Heparinase-III gen
2	250.5	7.2	1902	24	ABN67223	Streptococcus poly
C 3	250.5	7.2	2155561	24	ABN71527	Streptococcus poly
4	234.5	6.7	1923	24	ABN67224	Streptococcus poly
C 5	176	5.0	3763	19	AAV52319	Streptococcus pneu
C 6	166.5	4.8	11427	20	AAQ31402	Enterococcus faeca
7	139	4.0	2295	24	AAQ69654	Listeria innocua D
C 8	139	4.0	3011208	24	ABQ69245	Listeria innocua D
9	137.5	3.9	3495	24	ABN70606	Streptococcus poly
10	137.5	3.9	3522	24	ABN67343	Streptococcus poly
11	132.5	3.8	2097	24	ABN91332	Staphylococcus epi
12	127	3.6	3994	5	AAAN40009	Saccharomyces cere
13	127	3.6	3994	12	AAQ11883	Invertase gene. S
14	123.5	3.5	3414	14	AAQ41744	DNA encoding B.t.
15	123	3.5	1542	19	AAV09317	Unsecreted Yeast i
16	122.5	3.5	3753	21	AAAO5807	Group B Streptococ
17	121	3.5	3414	15	AAAO5821	Bacillus thuringie
18	120.5	3.4	3756	24	ABN67342	Streptococcus poly
19	120.5	3.4	3759	24	ABN70380	Streptococcus poly
C 20	120.5	3.4	4403765	22	AAI99683	Mycobacterium tube
C 21	120.5	3.4	4411529	22	AAI99682	Mycobacterium tube
22	118.5	3.4	3414	15	AAQ56922	Bacillus thuringie
23	118	3.4	2319	17	AAQ56926	Heparinase-II gene
24	118	3.4	7787	24	ABQ71070	Listeria monocytog
25	117.5	3.4	4455	18	AAAT25095	H. pylori secreted
26	117.5	3.4	8709	22	AAAF25628	H. pylori HPN165 e
27	117.5	3.4	2944528	24	ABA03041	Listeria monocytog
28	116	3.3	2215	21	AAQ43884	C. vicina LSP-2 cd
29	115.5	3.3	3414	14	AAQ41745	DNA encoding B.t.
30	115.5	3.3	3504	18	AAAT62498	Nematocidal toxin
31	115.5	3.3	1163020	24	ABQ67197	Listeria innocua c
32	115.5	3.3	3011208	24	ABQ69245	Listeria innocua D
33	115	3.3	3222	24	ABK75363	Bacillus lichenifo
34	115	3.3	16535	19	AAV52207	Streptococcus pneu
C 35	115	3.3	92584	21	AAAT2288	BAC containing rep
C 36	115	3.3	1830121	17	AAAT42063	Haemophilus Influe
37	114.5	3.3	2859	19	AAV42656	Human adenovirus 5
38	114.5	3.3	3504	18	AAAT89185	Nematode toxin 167
39	114.5	3.3	3504	18	AAAT61018	Bacillus thuringie
40	114.5	3.3	3903	19	AAK14480	H. pylori GPHO 109
41	114.5	3.3	4785	14	AAQ48797	B. ohbensis Trp(188
42	114.5	3.3	1830121	17	AAAT42063	Haemophilus Influe
43	113.5	3.2	3845	20	AAK59337	Avian topoisomeras
44	113.5	3.2	4765	17	AAAT08132	Protease gene. Py
45	113.5	3.2	4765	18	AAAT85670	Pyrococcus furiosu

ALIGNMENTS

1. **THESE**

RESOLUTION
AA099227

ID: AA099227 standard: DNA: 1980 BP.

XX
XX

AC AA099227;

XXXXXX

DT 07-APR-1996 (first

XX

DE Heparinase-III gene.

XXIV

KW Heparinase-III; hepa

KW *Escherichia coli*; po

KW toxicity; vector; ha

KW affinity chromatogra

XX

OS Flavobacterium hepar

```
sig_peptide      1..72      /*tag= a
mat_peptide      73..1977   /*tag= b
```

XX
PN WO9534635-A1.

XX
PD 21-DEC-1995.

XX 09-JUN-1995; 95WO-US07391. PF

XX
PR 10-JUN-1994: 94US-0258639

XX PA (TBEX-) TBEX TECHNOLOGIES

PA (ZIMM/) ZIMMERMANN J.
XX

PI
Bennett C, Blain F, Gu K, Musil R, Su H, Zimmermann J;
XX

DR WPI; 1996-097381/10.
DR P-DCOE. AAP89025

[illegible]

III - for degrading heparin and heparan sulphate, also related host cells, proteins and antibodies, useful in heparinase purificn.

PS Claim 8; Fig 8; 75pp; English.

The sequence encodes Flavobacterium heparinum heparinase-III (EC-4.2.2.8), which degrades heparan sulphate. The sequence is isolated by polymerase chain reaction amplification using e.g. primers 3-1 to 3-4 (AAQ992324-Q99237) or degenerate primers 3-5 to 3-8 (AAQ99238-Q99241). Toxicity of the gene and natural selection of the host against clones with the entire sequence has been circumvented by cloning sections of the gene separately in *Escherichia coli* using a vector with a modified ribosome binding site, which increases expression levels. The heparinase-III may be used to neutralise anticoagulant activity. Antibodies against the protein may be used to differentiate between native and recombinant enzymes, and when immobilised they may be used for heparinase purification by affinity chromatography.

Sequence 1980 BP; 629 A; 432 C; 439 G; 480 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	1980
Score:	3494.00	Matches:	659
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	17	Gaps:	0

US-09-802-285A-2 (1-659) x AAQ99227 (1-1980)

QY 1 MetThrLysIlePheIysArgIleIleValPheAlaValIleAlaLeuSerSerCly 20
 |||||
Dd 1 ATGACTACGAATAATTTTAAAGGATCATGTATTTGCTGTAATTGCCCTATCGTCGGGA 60

Qy 21 AsnIleLeuAlaGlnSerSerSerIleThrArgLysAspPheAspHisIleAsnLeuGlu 40
Db 61 AATATCTTGACAAAGCTCTTCGATTACCGAGAAAGATTTTGACCACTCAACCTTGAG 120

QY 41 TyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAspAlaAla 60
DB 121 TATTCCGGACTGGAAAAGGTTATAAAGCAGTTGCTGCCGGCAACTATGACGATGCGGCC 180

QY 61 LysAlaLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsn 80
| | | | |
Dd 181 AAAGCATTACTGGCATACCTACAGGGAAAAGAATAAGGCCAGGGAACTGTATTCAGTAAT 240
| | | | |

QY 81 AlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAsp 100
 D6 241 GCAGAAAGCCCTGCCGATATACGCCAGCCATAGTAAGTTACGCTGAATATGCTCGAC 300

Qy	101	LysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLys	120
Db	301	AAGCGTTTGGTCCACAGTTTCAACCGCACAAAGGCTACGGCTATTTTGATTATGGTAA	360
Qy	121	AspIleAsnTrpGlnMetTrpProValLysAspGlnValArgTyrPheGlnLeuHisArg	140
Db	361	GACATCAACTGGCAGATGTGGCGGTAAAGACAAATGAAGTACGCTGGCAGTGGCACCGT	420
Qy	141	ValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla	160
Db	421	GTAAAAATGGTGGCAGGCTATGGCCCTGGTTTATCACGCTACGGGCGATGAAAAATATGCA	480
Qy	161	ArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGln	180
Db	481	AGAGAAATGGGTATATCATGACAGCGATTGGGCCAGAAAAACCCTATGGGCGCTGTGCGAC	540
Qy	181	AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro	200
Db	541	GATAATGATAAAATTTGTGTGGCGCCCTTCGAGTGTGGACAGGTTACAAAGTCTTCC	600
Qy	201	ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe	220
Db	601	CCAACTTCAGCTTATTGTAAACTCGCACGCTTATCCCGACGCTTTTAAATGGAATTT	660
Qy	221	LeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsn	240
Db	661	TTAAACAGTTACCAACCAACAGCCGATTAATTTACTCGCAATATGCCGAAACAGGGAAC	720
Qy	241	HisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys	260
Db	721	CACCGTTTATTGAAGCCCAACGCAACTGTGTTTCAGGGGTATCTTTCCCTGAAATTTAA	780
Qy	261	AspSerProArgTrpArgGlnThrGlyLeuSerValLeuAsnThrGluLeuLysGln	280
Db	781	GATTCAACAGATGGAGCAAAACCGGCATATCGGTGCTGAACCGAGATCAAAAACAG	840
Qy	281	ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAsp	300
Db	841	GTTTATGCCGATGGATGCAGTTTGAACCTTTCACCAATTTACCATGTAGCTGCCATCAT	900
Qy	301	IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln	320
Db	901	ATCTTCTTAAAGGCTATGTTCTGCAAAACAGATTAACTTTGAAAAGAAATTTCCGCAA	960
Qy	321	SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp	340
Db	961	TCTTATGTACAACTGTAGAAATATGATATATGGCGCTGATCAGTATTTTCACTGCCAGAT	1020
Qy	341	TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGln	360
Db	1021	TATAACACCCCTATGTTTGGAGATTCATGATTACAGATAAAAAATTCAGGATGGCACAG	1080
Qy	361	PheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAsp	380
Db	1081	TTTGCCAGCTGGCCCGGGTTTTCCCGGCAACCGGCCATAAAATATTTTGTACACAGAT	1140
Qy	381	GlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyr	400
Db	1141	GGCAAAACAGGTAAAGGCGCTAACTTTTATCCAAAGCATTTGAGCAATGACAGGCTTTTAT	1200
Qy	401	ThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProPro	420
Db	1201	ACGTTTTAGAAGCGGATGGGTAAATAATGCAACCGTTATGGTATTAAAGGCGACGCTCC	1260
Qy	421	GlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsn	440
Db	1261	GGGGAAATTCATGCCACCGCGGATAACGGGACTTTTGAACTTTTATAAAGGGCGAGAAAC	1320
Qy	441	PheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArg	460
Db	1321	TTTACCCCAACAGCCCGGGTATTGTGTATAGCGGCAACGAAGCCATCATGAAATCGCGG	1380
Qy	461	AsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetVal	480


```

Db 1381 AACTGGTACGTCACACCCGATACACAGACGCTTACATCGACATCAAAATATGGTC 1440
Qy 481 ILeThrIysAlaArgGlnAsnLysTyrGluThrGlyAsnAsnLeuAspValLeuThrTyr 500
Db 1441 ATTACCAAGCCGGCAAAACAAATGGGAAACAGGAATTAACCTTGATGTGCTTACCTAT 1500
Qy 501 ThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheLeuAsnLysLys 520
Db 1501 ACCAACCCAGCTATCCGAATCTGGACATCGACGCACTGCTACTTTTCATCAACAAAAA 1560
Qy 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTyr 540
Db 1561 TACTTTCGTCTCGATAGGGCAATAGGCGAAGCTACCGGAACCTGGCGGTACACTGG 1620
Qy 541 GlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThr 560
Db 1621 CAGCTTAAGAAGACAGCAACCCCTGTTTCGATAAGACAAAGAACCGGGTTTACCCACT 1680
Qy 561 TyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn 580
Db 1681 TACAGAGATGGTAAACACCTGATCCATCGTGAATCGGACAGGACGAGCCTCAAT 1740
Qy 581 GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal 600
Db 1741 GAAGACAGAGAAAGGTATCTTATGTTTACAATAAGGAGCTGAAAAGACCTGCTTCGTA 1800
Qy 601 PheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyr 620
Db 1801 TTTGAAAGCCTAAAAAGAAATGCGGCACACAAAATTTTGTCTAGTATAGTTTATCCATAC 1860
Qy 621 AspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLys 640
Db 1861 GACGGCCAGAGGCTCAGAGATCAGATACGCGGAAACCAAGGCAATGTTTGGAAA 1920
Qy 641 GlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnGlnLeuValLeuValPro 659
Db 1921 GGCAAGCTTAATCTAACCTTACCATTAAACGGAACAAACAGCTTGTTGGTTCT 1977

RESULT 2
ID ABN67223
ID ABN67223 standard; DNA; 1902 BP.
AC AC
AC ABN67223;
XX
XX 01-JUL-2002 (first entry)
XX Streptococcus polynucleotide SEQ ID NO 2359.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
XX Streptococcus agalactiae.
XX
XX W0200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX 24-NOV-2000; 2000GB-0028727.
XX 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
XX Tettelin H;
XX
XX WPI; 2002-352536/38.
XX
XX P-PSDB; ABP26592.

```

```

XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
XX for detecting a compound that binds to the protein -
XX
PS Claim 7; Page 3387-3388; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 1902 BP; 641 A; 264 C; 362 G; 635 T; 0 other;

```

```

Alignment Scores:
Pred. No.: 1.82e-14 Length: 1902
Score: 250.50 Matches: 127
Percent Similarity: 36.09% Conservative: 74
Best Local Similarity: 22.80% Mismatches: 223
Query Match: 7.17% Indels: 133
DB: 24 Gaps: 21

US-09-802-285A-2 (1-659) x ABN67223 (1-1902)
Qy 123 AsnTyrGlnMet-----Tyr-----Tyr----- 127
Db 145 AATTGGGATATGGAACCTTCTTAAGCATATGTCTTTCATCTTGGATGGGATAAG 204
Qy 128 ProValLysAspAsn--GluValArgTyrGlnLeuHisArgValLysTyrTyrGlnAla 146
Db 205 CCAGTACAGATGATCCAGATGGTGTATATGCTAAATCGTCAAACTTATCTTTTAAA 264
Qy 147 MetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTyrValTyrGln 166
Db 265 TTCTGTGTGCTTACATTGTTGAAGCGATAAGTCTTACCTTAGACAAATATATTTT 324
Qy 167 TyrSerAspTrpAla-----ArgLysAsnProLeuGlyLeuSerGlnAspAsn 182
Db 325 ATGTACCATTTGGATTGATTGCTCAATTTACACTAAGCCAGAGGGCGG----- 372
Qy 183 AspLysPheValTyrArgProLeuGluValSerAspArgValGlnSerLeuProProThr 202
Db 373 -----GTAAGTCGTACTATTGACACTGGGATAGCTTGCATGCTCAAGTT 423
Qy 203 PheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsn 222
Db 424 TTGATTTTTTTAGATTATTTTGATTAAATAACAGAACTAAAAAATAAACTATTAACT 483
Qy 223 SerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln-----Gly 239
Db 484 AGTTTACGGGAGCAGATAACTTATATCGAGGAGCATATATCGTGAAGAAGATAGTCTAAGT 543
Qy 240 AsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPhe 259
Db 544 AACTGGGAATTTTACAAACACAGCAATATTTGGCGTGTATTACTATTATGAGATGAA 603
Qy 260 LysAspSerProArgTyrArgGlnThrGlyIleSerValLeuAsnThrGluIleLysLys 279
Db 604 TTAATCTACCTCAAAATTCAGAGCTTTTGTCTGAAGAAGATATTATTACTTCAAACTCAAGCTT 663

```

```
QY 280 GlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIle 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
664 CAAATTTAGATGACGGGACGATGATGACATCAATATGATGATGATGATGATGATGATG 723
QY 300 AspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPhePro 319
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
724 AAGTCCTTAATGAAGTAGTATCTCTCGCCCTAAATATATCTACCATTAGAA----- 777
QY 320 GlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuPro 339
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
778 -----GAAACTATTGAAAAATGGTTTACCTATCTAATCTGATGATGATGATGATG 925
QY 340 AspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAla 359
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
826 GATTATGTCAACTAGCTATAGGGACAGTGTGACCGATACTCGT----- 873
QY 360 GlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThr 379
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
874 GATATTCTAATCTGGCAACACTGTGTGAAATCTTCTAAACAAAAATCAITTTCTTTT 933
QY 380 AspGlyLysGln-----GlyLysAlaProAsnPheLeuSerLys 392
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
934 GATAATGTTAATTAGAACTTTACTTTTGTGGGAGCCATCAATTTATCTTTTGTAA 993
QY 393 AlaLeuSerAsnAla-----GlyPheTyrThrPheArgSerGlyTrpAsp--- 407
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
994 GAAATACCGCGTGGCAACATAGGAGAGTCTGCTTATCTTTTCCAGATCTGTCATG 1053
QY 408 -----LysAsnAlaThrValMetValLeuLysAlaSerProProGlyGluPheHis 424
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1054 TGTTTACGTGATGATAGCGCTTATATATTTTAAAGAACGGTCCATTTGTCAGCGTCAT 1113
QY 425 AlaGlnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAsp 444
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1114 ACTCATGTGATGATATAGTGTGCTCTATGATATAAAAGAACCTATTTTCATGAT 1173
QY 445 AlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArg 464
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1174 GCAGGAAGATATACCTTACAAAGAAAGAACAA-----CTAAGGTATGATTTTAA 1221
QY 465 GlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAla 484
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1222 CGTTCGATGATGATCAATCAATGATACCTTGATGGCAACCCCTTGAAGATGATCAAG--- 1278
QY 485 ArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnProSer 504
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1279 -----GACTCTGGACATACAT-----TCT 1299
QY 505 TyrProAsnLeuAspHis----- 510
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1300 TATCCAAAATGTGACTATTGTGCTGACTTCAAGGATAGGTACCATTTAGTCGAAGGA 1359
QY 511 -----GlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIle 525
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1360 CAACTACATGTCGAAGAGCTCTGATATCTATTACCATAGCGATGGTGTGTAACCTTA 1419
QY 526 AspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAsp 545
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1420 CCGCAGGCAT----- 1431
QY 546 SerAsnProValPheAspLysThrLys-----AsnArgValTyrThrThr----- 560
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1432 ---ACCTTAGTTATTGATAGGTAGTGTGTCAGGAGAGCATGCTTAACAAATCAATAT 1488
QY 561 -----TyrArgAspGly-----AsnAsnLeuMetIleGlnSer 571
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1489 ATTTTAGATGATCAGGTCAATTAAGAAATGGTTGTTAATGACATGAAATAGTAAGT 1548
QY 572 LeuAlaAspArgThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsn 591
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1549 -----CCTACGACCTTATCTAGAGATTGCTTATTTCTAAGCGGTATAT 1596
QY 592 -----LysGluLeuLysArgProAlaPheValPheGlu 602
```

Db 1597 CAATTGACAGAAAGTCATAAATAGTTAGAAAATAAAATTTGTTGATGAG 1647

RESULT 3

ABN71527/c

ID ABN71527 standard; DNA; 2155561 BP.

XX AC ABN71527;

XX DT 02-JUL-2002 (first entry)

XX DE Streptococcus polynucleotide SEQ ID NO 10967.

XX KW Streptococcus; GAS; CBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
XX KW antinflammatory; infection; vaccine; meningitis; gene therapy; ds.

XX OS Streptococcus sp.

XX PN WO200234771-A2.

XX PD 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB04789.

XX PR 27-OCT-2000; 2000GB-0026333.

XX PR 24-NOV-2000; 2000GB-0028727.

XX PR 07-MAR-2001; 2001GB-0005640.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;

XX PI Tettelin H;

XX DR WPI; 2002-352536/38.

XX PT New Streptococcus protein for the treatment or prevention of infection

XX PT or disease caused by Streptococcus bacteria, such as meningitis, and

XX PT for detecting a compound that binds to the protein -

XX PS Claim 8; Page 4196-4488; 4525pp; English.

XX CC The invention relates to a protein (ABP25413-ABP30895) from group B

XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

XX CC the specification. The proteins have antibacterial and antinflammatory

XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and

XX CC antibodies that bind (I) are used in the manufacture of medicaments for

XX CC the treatment or prevention of infection or disease caused by

XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a

XX CC biological sample. (I) is used to determine whether a compound binds to

XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

XX CC used as a vaccine or diagnostic composition. The disease caused by

XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic

XX CC acid encoding (I) may be used to recombinantly produce (I) and may be

XX CC used in gene therapy. Antibodies to (I) are used for affinity

XX CC chromatography, immunoassays, and distinguishing/identifying

XX CC Streptococcus proteins.

SQ Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;

Alignment Scores:

Pred. No.: 2,79e-10 Length: 2155561

Score: 250.50 Matches: 127

Percent Similarity: 36.09% Conservative: 74

Best Local Similarity: 22.80% Mismatches: 223

Query Match: 7.17% Indels: 133

DB: 24 Gaps: 21

US-09-802-285A-2 (1-659) x ABN71527 (1-2155561)

```

QY 123 AsnTrpGlnMet-----Trr-----127
Db 1845827 AATTGGGATAGGACCTGTTCTAAAGCATATGTCTTGATCCTTTGGAGTGGGATAAG 1845768
QY 128 ProValIysAspAsn---GluValArgTrpGlnLeuHisArgValIysTrpTrpGlnAla 146
Db 1845767 CCAGTAACAGATGATCAGAAATGGTTGTATATGCTAAATCGTCAAACTATCTTTTAAA 1845708
QY 147 MetAlaLeuValTyrHisAlaThrGlyAspGluIysTyrAlaArgGluTrpValTyrGln 166
Db 1845707 TTCTGTGTGCTACATTTGTGAAGGCGATAAAGTCTTACCTTAGACAAATGAATATTTT 1845648
QY 167 TyrSerAspTrpAla-----ArgIysAsnProLeuGlyLeuSerGlnAspAsn 182
Db 1845647 ATGTACCATTTGGATTGTTCAATTTACACTAAAGCCAGAGGGCG-----1845600
QY 183 AspIysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProProThr 202
Db 1845599 -----GTAAGTCGTACTATTGACACTGGGATACGTTGCATGTCATGGCTCAAAGTT 1845549
QY 203 PheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsn 222
Db 1845548 TTGATTTTTTTAGATATTTTGGATTAAATAACAGAACTAAAAAAATTAACCTATTAACT 1845489
QY 223 SerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln-----Gly 239
Db 1845488 AGTTACGGGACGACATACTTATATGAGGACTATTATCTGTGAAAAAGATAGTCTAAGT 1845429
QY 240 AsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPhe 259
Db 1845428 AACTGGGGAATTTTACAAACAACAGCAATATTGCGGTGTTTATACTATTATGAAGATGAA 1845369
QY 260 LysAspSerProArgTrpArgGlnThrGlyLeuSerValLeuAsnThrGluIleLysLys 279
Db 1845368 TTTAACTCTAGTAAATTCAGAGTTTTCGTAAGAGAAATTTATTTACTTCAATCAAGCTT 1845309
QY 280 GlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIle 299
Db 1845308 CAAATTTTAGATGACGGGACCCAGTATGACAACTAATATGATCATGTAGAGTCTTG 1845249
QY 300 AspIlePheLeuIysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPhePro 319
Db 1845248 AAGTCCTTAAATGGAACATGATGATCTCGCGCTAAATATATCTACCATTAGAA-----1845195
QY 320 GlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuPro 339
Db 1845194 -----GAACTATTTGAAAAATGTTATCTCGCTAAATATATCTACCATTAGAA-----1845147
QY 340 AspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspIysAsnPheArgMetAla 359
Db 1845146 GATTATTTGCACTAGCTATAGGGACAGTCAATGACCGACTACTCGT-----1845099
QY 360 GlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThr 379
Db 1845098 GATATTCTAATCTTGGCAACACTGTGTTGTAATCTCTTAAAAACAAAATCATTTCTTTT 1845039
QY 380 AspGlyLysGln-----GlyLysAlaProAsnPheLeuSerLys 392
Db 1845038 GATAAGTTAATTAGAACTTTACTTTGTTGGGAAGCCATCAATTTATCTTTTGA 1844979
QY 393 AlaLeuSerAsnAla-----GlyPheTyrThrPheArgSerGlyTyrAsp---407
Db 1844978 GAAATACCGCGTGGACAAATAGGAGTCTGCTATCTTTTCCAGATTCGTGTCATGTG 1844919
QY 408 -----LysAsnAlaThrValMetValLeuIysAlaSerProGlyGluPheHis 424
Db 1844918 TGTTTACGTGATGATAGCGGTATATATTTTAAAGAACGTCATTCATTTGGTAGCGTCTAT 1844859
QY 425 AlaGlnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAsp 444
Db 1844858 ACTCATAGTATATAATATGTTGTTCTCTGATATAAAGAACCTATTTTCAITGAT 1844799
QY 445 AlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTyrTrpArg 464

```

Db 1844798 GCAGGAGATATATCTTACAAGAGAACAA-----CTAAGTATGATTTTAAA 1844751

QY 465 GlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAla 484

Db 1844750 CGTTCGACTAGTCATTCAACATGTCACCTTGATGGGCAACCCCTTAGAAATGATCAAG-- 1844694

QY 485 ArgGlnAsnIysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnProSer 504

Db 1844693 -----GACTCTTTGGACATACAAT-----TCT 1844673

QY 505 TyrProAsnLeuAspHis-----510

Db 1844672 TATCCAAAATGTGACTATTTGTCAGTTGCTTCAAGATAGGTACCATTTAGTCGAAGGA 1844613

QY 511 -----GlnArgSerValLeuPheIleAsnLysIysTyrPheLeuValIle 525

Db 1844612 CAACTACATGTCCAAAGAGCTTCTGATATCTATTACCATAAGCGATGTTGTTAACTTTA 1844553

QY 526 AspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAsp 545

Db 1844552 CCGAGGCCATT-----1844541

QY 546 SerAsnProValPheAspLysThrLys-----AsnArgValTyrThr-- 560

Db 1844540 ---ACCTTAGTTATGATAGGTGAGTTGCCAGGAGAGCATGCTTAAACAATCAATAT 1844484

QY 561 -----TyrArgAspGly-----AsnAsnLeuMetIleGlnSer 571

Db 1844483 ATTTTAGATGATCAGGTCATTTATGAAATGGGTTTGTAAATGACTTGAATTAGTAAGT 1844424

QY 572 LeuAsnAlaAspArgThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsn 591

Db 1844423 -----CCTAGCAGCTTTAATCTAGAAGATTGCCCTTATTTCTAAGCGGTATAAT 1844376

QY 592 -----LysGluLeuLysArgProAlaPheValPheGlu 602

Db 1844375 CAATTGCAGAAAGTCAATAAATTAGTTAGAAAAATAAAATTTGTTGATGAG 1844325

RESULT 4

ABN67224

ID ABN67224 standard; DNA; 1923 BP.

XX ABN67224;

AC ABN67224;

XX 01-JUL-2002 (first entry)

DT XX

DE Streptococcus polynucleotide SEQ ID NO 2361.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;

KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.

OS Streptococcus pyogenes.

XX WO200234771-A2.

PN 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

PF 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

PI Tettelin H;

XX WPI; 2002-352536/38.

DR P-PSDB; ABP26593.


```
QY 355 AsnPheArgMetAlaGlnPheAla----- 362
Db 1378 ACAGAAATTTGAGCTGTCTGCTGGGTTTGAACACGGAAGACCTTCTTAACGGTCTG 1319
QY 363 -----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPhe 377
Db 1318 GATGTTAAAGTTGATTTGCTTAGCTCTTCTTCTGGCGGAGAAAGGTCAAGCGACTG 1259
QY 378 AlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAla 397
Db 1258 CAGGAATTTCAAAAGAGAGCTTGGCAGCCTAAG-----TCCATGATCTTTGAAGACTCT 1205
QY 398 GlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAla 417
Db 1204 GGACATGCTGCATTAAGAT-----GAACATCGTATCTATTCTTTCAAAAT 1157
QY 418 SerProProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheLys 437
Db 1156 GGTCCGCTAGGAAGTGCCTATAGTCATAGCAGAGAATAGTTTTTGTCTTACAGTATCAA 1097
QY 438 GlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMet 457
Db 1096 GGCCAACTTATTTTCATAGATGCTGGCGCTTATCTTATCG-GGAGATATATCAACGTTA 1038
QY 458 LysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGln 477
Db 1037 TCTCTTAAAGAGTCTTGGAGTCATTCGACCTGCATGT----- 999
QY 478 AsnMetValIleThrLysAlaArgGlnAsnLysTrpGlu---ThrGlyAsnAsnLeuAsp 496
Db 998 -----AGATGGAAAGTCCCGGAAAGAATCACGGG 969
QY 497 ValLeu-ThrTyr----- 500
Db 968 ATCTCGGAATATGAATATCTCTCACTCCCTGTTTGTCCACCATAAAGAAAGGAGGG 909
QY 501 -----ThrAsnProSerTyrProAsnLeuAspHi 510
Db 908 AGTGCATATATAGGGGGCTTATTTGGTCAGCAGAACCTGATTTGCCTTATCTT---CA 852
QY 510 sGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAsp-----ArgAl 528
Db 851 CAGGAGAAATATCTCATGTTGTAGAGGATGCTGGCTCTTGGTAGATCATCATCAGGTTG 792
QY 528 aileGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnPr 548
Db 791 TCAAGGTCAA-----CATGAGGTGTTGACTCAGTTT----- 761
QY 548 oValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGly-----AsnAs 566
Db 760 -ATCCTTGACAGGAT-----GTGACCTATCAAGATGGGAAATCATCA 717
QY 566 nLeuMetIleGlnSer 571
Db 716 GTTGAGACTATGGAGT 701
```

RESULT 6

```
AA13102/c
ID AAX13102 standard; DNA; 11427 BP.
XX
AC AAX13102;
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:165.
XX
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.
XX
PN W09850555-A2.
```

```
XX 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US08985.
XX
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Dillon PJ, Kunsch CA;
XX
XX WPI; 1999-045171/04.
XX
DR New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
PS Claim 1; Page 932-938; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SQ Sequence 11427 BP; 3451 A; 2545 C; 1839 G; 3587 T; 5 other;
```

Alignment Scores:

```
Pred. No.: 3,378-05 Length: 11427
Score: 166.50 Matches: 121
Percent Similarity: 35.18% Conservative: 89
Best Local Similarity: 20.27% Mismatches: 215
Query Match: 4.77% Indels: 174
DB: 20 Gaps: 26
```

US-09-802-285A-2 (1-659) x AAX13102 (1-11427)

```
QY 120 LysAspIleAsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHis 139
Db 7871 AAAGATATGCTGGTGAATCGGTATCTGATGATGATCCAGATGGCTCTTTATGTAGT 7812
QY 140 ArgValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyr 159
Db 7811 CGTCAAGCTTCTTGTAGATCTGGCACAAGCATATGCATTTACTTAAAAAAGAACGTTAC 7752
QY 160 AlaArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSer 179
Db 7751 TTACAGAAATGGCAGAGCTTGGCTGCTTAGATGTTGGATTCGATGATCAAAAC--- 7641
QY 180 GlnAspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeu 199
Db 7697 AATTCGACGATAGGAGTGTTCGGCTCGTGTAGATGTTGGATTCGATGATCAAAAC--- 7641
QY 200 ProProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGlu 219
Db 7640 -----TGGATGAAAGAGCTTGAAGCTATATTTCCAAATCGCTGATTTTCA 7599
QY 220 PheLeu-----AsnSerTyrHisGlnGlnAlaAspTyrLeuSer 232
Db 7598 CTATTAGGAATTGATGATGTGTTGAACACACGCTTGTCTGATCATCTGACTATTGGAG 7539
```


QY 523 LeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeu 542
 Db 1903 -----GCAAAAGATAATTTAAACTTTCAGGAGCA 1935
 QY 543 LysGluAspGlnProValPheAspLysThrLysAsnArgValTyrThrTyrArg 562
 Db 1936 AAATGGGATAGT-----GAAATTTCTCAAGTAGAGAA 1968
 QY 563 AspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn----- 580
 Db 1969 AAAGGTCAAAAAGTAACATAATTAATTCG-----AAAACAAGTATAAATGGTTGG 2016
 QY 581 -----GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgPro 597
 Db 2017 CATGAAGTAACCTACGCGTGTAAATAAGGTTATAG-----ATTTTAAGTAT 2064
 QY 598 AlaPheValPheGluLysPro 604
 Db 2065 AACTATTATTAGTCGAAACCA 2085
 RESULT 8
 ABQ69245/c
 ID ABQ69245 standard; DNA; 3011208 BP.
 XX
 AC ABQ69245;
 XX
 DT 29-AUG-2002 (first entry)
 XX
 DE Listeria innocua DNA sequence #694.
 XX
 KW Antibacterial; Listeria; food contamination; mutational analysis;
 XX infection; ds.
 XX
 OS Listeria innocua.
 XX
 PN W0200228891-A2.
 XX
 XX 11-APR-2002.
 XX
 XX 04-OCT-2001; 2001WO-FR03061.
 XX
 XX 04-OCT-2000; 2000FR-0012697.
 XX
 XX (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 XX Kunst F, Glaser P;
 PI WPI; 2002-332479/37.
 XX
 XX New genomic sequences from Listeria species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators -
 XX
 XX Claim 5; SEQ ID 2058; 180pp; French.
 XX
 XX The present invention relates to nucleic acid sequences
 CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
 CC and primers for identification and/or detection of Listeria (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of Listeria (potential therapeutic agents), also for
 CC treating infections by Listeria, and are useful as immunogens in
 CC anti-Listeria vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 3011208 BP; 941651 A; 568176 C; 559189 G; 942192 T; 0 other;
 Alignment Scores:

Pred. No.: 33.9 Length: 3011208
 Score: 139.00 Matches: 148
 Percent Similarity: 31.94% Conservative: 97
 Best Local Similarity: 19.30% Mismatches: 280
 Query Match: 3.98% Indels: 242
 DB: 24 Gaps: 37
 US-09-802-285A-2 (1-659) x ABQ69245 (1-3011208)
 QY 7 LysArgIleIle-----ValPheAlaValIle-----AlaLeuSerSerGly 20
 Db 2117803 AAAAGAAATTATATCTGTATTTGTCATTTTCTACTGGGATTTTAGCCCATTC 2117744
 QY 21 AsnIleLeuAlaGlnSerSerSerIleThrArgLysAspPheHisIleAsnLeuGlu 40
 Db 2117743 ACAATTAAGCAAGTGCAGCAACCAATACAAATAAATTTGAAGAGCTT----- 2117693
 QY 41 TyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAspAlaAla 60
 Db 2117692 -----TATGCAAGAGCTAAA 2117678
 QY 61 LysAlaLeu---LeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSer 79
 Db 2117677 AAACATTTAGGTAAACCATATTCGCAAGACAGGACAGAGAGAGGCGCCAAATTTACTTC 2117618
 QY 80 AsnAla-----GluLysProAlaAspIleArgGlnPro--- 90
 Db 2117617 GACTGCTCAGGTTACACACAATATGTTTATGAATAAGTAAACAGGAGTAAGAATTCCTAAT 2117558
 QY 91 -----IleAspLysValThrArgGluMetAlaAspLysAla 102
 Db 2117557 ACTTCAGCTCCTCAATATTCAGCTGCTGATTAAGTGAAATGGTAATCAAAAACCTGGT 2117498
 QY 103 LeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLysAspIle 122
 Db 2117497 GATTTAGTATATTTCAAGGCTCATGTAGGAATA---TACATAGTAAATGCTAAATGATT 2117441
 QY 123 AsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArgValLys 142
 Db 2117440 AATGCTCAAAATGATGAGTCAAAATAGACAACATT-----AATAGTAGT 2117396
 QY 143 TrpTrpGlnAlaMetAlaLeuValTyr-----HisAlaThrGlyAspGluLysTyrAla 160
 Db 2117395 TATTCGAGAGTATTTTGTGTGATGCGAAGATTTTCAATTTCTCCGAGAAAAAAGGA 2117336
 QY 161 ArgGluTrpValTyrGlnTyrSerAsp-----Trp----- 170
 Db 2117335 TCTAAATCCGCTATGCTGTGTCAGATTTAAATTTAGATCTAGCAATAACTGGATAGT 2117276
 QY 171 -----AlaArgLysAsnProLeuGly-----LeuSerGlnAspAsnAspLys 184
 Db 2117275 TCAGTAGCTGGAAGATTCACCAAGGCGCTAAAGTATCTATTGACTTACAGTATGAAG 2117216
 QY 185 PheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProProThrPheSer 204
 Db 2117215 AATGCTTGGTGTATGCTTAATATAATACTAAAGTTATATGTTAAATACACTAAT 2117156
 QY 205 LeuPheValAsnSerProAlaPheThrProAlaPhe----- 216
 Db 2117155 TATTTCTCCGATACACCTGTGATAAAAAACATCATATGCCAAGGATATATAAACTTAAGA 2117096
 QY 216 ----- 216
 Db 2117095 ACGAAAGCAACATGGGATAGCGATGTAGCTCAAAAAGTACAAAAGGTGAAAAGTTACT 2117036
 QY 217 -----LeuMetGluPheLeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHis 234
 Db 2117035 GTGAACCTTAAACTAATGTGATGTTGTTGTTCAAGTACATACACGCGGAAAAAAGCAGC 2116976
 QY 235 TyrAlaGluGlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeu-----PheAla 252
 Db 2116975 TATATGATTTTAAATAATAATTACTTTAGTAGAAAATCCCTTGAATATGGAACGTATTAT 2116916

QY 253 GlyValSerPheProGluPheLysAspSerProArgTyrArgGlnThrGlyIleSerVal 272
Db 2116915 GCGGTTGGTACCTTAAATCGGTAGTCAGCAACTGG---CATAGTAGCATTTAGCTTA 2116959
QY 273 LeuAsnThrGlu-----lLeLysLysGlnValTyrAlaAspGlyMetGlnPheGlu 289
Db 2116858 GTTGTACCAAGAGCGAGCTGTAAAAGTAGAA-----ATGGATACAAAT 2116914
QY 290 LeuSerProIleTyrHisValala----- 297
Db 2116813 AGTGGTCCCTTGGTATTAAGTAAACATATCAAAACCAACAGGTTCATCCCTCAACAGAT 2116754
QY 298 -----AlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu 314
Db 2116753 GATTATTTATCCAAACTACTGTTCTAAAACGTTATACGCTTAAGATATATTAACCTTA 2116694
QY 315 GluLysGlu-----PheProGlnSerTyrValGlnThrValGlu-----AsnMetIle 330
Db 2116693 AGAACCAAAAGCAACTTTGGGATAGCGAGCTTGCTCAAAAAGTACAAAAGGTGAAAAGTA 2116634
QY 331 MetAlaLeuIleSerIleSerLeuProAspTyrAsnThrProMetPheGly----- 347
Db 2116633 ACAGTTAAATTAAACCAAGTGTAAATGGTGTATCAAGTAACTTATGGTGTGTAACAAA 2116574
QY 348 -----AspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPhe 361
Db 2116573 GGTATACATCTTAATGATATACCTATTAGTCGAAAAGCGCTTAATATGAAAACCTAT 2116514
QY 362 -----AlaSerTrp-----AlaArg 366
Db 2116513 TATCGGTGAGTAGCTTAAATTTACGTAGTGGCGGAATGGGATAGTAGTAACTCA 2116454
QY 367 ValPheProAlaGlnAlaIleLys----- 377
Db 2116453 GTGGTTCAGAGGTAGAGCTGCAAAAGTCGAAATGACACGCAAGTGTGGCAATGGTTT 2116394
QY 378 AlaThr-----AspGlyLysGlnGlyLysAlaPro-----AsnPheLeuSerLys 392
Db 2116393 AAAGTAACCTATGACATATAACAGGTTATATGCCCTAAATGATTTATTTATCTGAA 2116334
QY 393 AlaLeuSerAsnAlaGlyPheTyrThr-----PheArgSerGly 405
Db 2116333 ACCGCTGTGTAAAAACCTATTATGCAAAAGATAATTTAACTTACGTAGCGAAGCAAAA 2116274
QY 406 TrpAspLysAsnAlaThrValMetValLeuLysAlaSerProProGlyGluPheHisAla 425
Db 2116273 TGGGATGACGAGTTACTCAAAAGTAGAAAAGGTGAAAAGTAAACAGTCAATTCGAA 2116214
QY 426 GlnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAla 445
Db 2116213 ACAAGTATCGATGGCTGGTATGAATG----- 2116187
QY 446 GlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGln 465
Db 2116186 -----ACATACGGTGGTAAAGAGCGCTATATGATTTTAAATAATAATATAT----- 2116142
QY 466 ThrArgIleHisSerThrLeuAspAsnGlnAsnMetVal-----lLeThrLys 483
Db 2116141 -----CTTGTAGCAGAACCTAGATTGTAACATATATACGAGTTAACACATTTAAATTTA 2116085
QY 484 AlaArgGlnAsnLysTrpGluThrGly----- 492
Db 2116084 CGTAGTGAATCTAAATGGGACAGCAGCATAGCAAGCAAGTGTACCTGAAAGCGCTAAAGTA 2116025
QY 493 -----AsnAsnLeuAspValLeuThrTyrThrAsnProSer 504
Db 2116024 AAAGTTGAATGAAACCAAGTGTAAATGGTATGTTACAAAGTAACTTATCAAAATAAACA 2115965
QY 505 -----TyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysTyrPhe 522
Db 2115964 GGTATATGCGGCTAAATGATTTATATATCTGAAAGTCTGTGTGTAACAACTTATAT 2115905
QY 523 LeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeu 542

Db 2115904 -----GCAAGATAATTTAACTTACGTAGCGAAGCA 2115872
QY 543 LysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrTyrArg 562
Db 2115871 AAATGGGATAGT-----GAAATTTCTCAAGTAGTAGAA 2115839
QY 563 AspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn----- 580
Db 2115838 AAAGGTGAAAAGTAACTATTATTCG-----AAACAAGTATAAATGTTGG 2115791
QY 581 -----GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgPro 597
Db 2115790 CATGAAGTAACTTACGGTGTGTAAGAAAGGTATATG-----ATTTAAGTGAT 2115743
QY 598 AlaPheValPheGluLysPro 604
Db 2115742 AACTATTATAGTCGAAAACCA 2115722
RESULT 9
ABN70606
ID ABN70606 standard; DNA; 3495 BP.
XX
AC ABN70606;
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 9125.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus pyogenes.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
PA (CHIR-) CHIRON SPA.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
DR WPI: 2002-352536/38.
DR P-PSDE; ABP29975.
XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
PS Claim 7; Page 4034; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.

XX
SQ Sequence 3495 BP; 1097 A; 733 C; 771 G; 894 T; 0 other;

Alignment Scores:

Pred. No.: 0.0045 Length: 3495
Score: 137.50 Matches: 152
Percent Similarity: 33.05% Conservative: 85
Best Local Similarity: 21.20% Mismatches: 263
Query Match: 3.94% Indels: 217
DB: 24 Gaps: 39

US-09-802-285A-2 (1-659) x ABN70606 (1-3495)

```
QY 43 GlyLeuGluLysVal-----AsnLysAlaValAlaAlaGlyAsn--- 55
Db 1453 GGTCAAGAAAAGTCATGGTTTGGATCCTTACGCCAAATCTCTCGCTGCTGGAATGAT 1512
QY 56 -----TyrAspAlaAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSerLys--- 72
Db 1513 GCGACTCTACTGATGACATCAAAACAGCAAAAGCTGCTTTATTGATCCAAAGCAAACTA 1572
QY 73 -----AlaArgGluProAspPheSerAsnAlaGluLysProAla 85
Db 1573 GGACCAACAGCGCTGATTGTCGCAAAATTAACAACTTTAAAGCGTGAAGCGCTATT 1632
QY 86 AspIleArgGlnProIleAspLysValThrArgGluMetAla---AspLysAlaLeuVal 104
Db 1633 ATCTATGAAGCACATGCGCAGATTTTACGTCAGATAAGCTCTAGAAGCGCAAGTTAACA 1692
QY 105 HisGlnPheGlnProHisLysGlyTyr---GlyTyrPheAspTyrGlyLysAspIle--- 122
Db 1693 CACCCCTTTGGGACTTTTTCAGCTTTGCTGGTGAACAGCTAGACTATCTCAAAAGACTTGGGG 1752
QY 123 -----AsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArg 140
Db 1753 GTTACCACAGCTTCATGTGCTACCGGTT----- 1779
QY 141 ValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla 160
Db 1780 -----TTGAGTTATTTTATCCCAATGAGCTGGCAAGAGC 1815
QY 161 ArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGln 180
Db 1816 CGCTCAACAGCCTACACGCTCTTCAGAC----- 1842
QY 181 AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro 200
Db 1843 AATAATTAACACTGGGGTTATGACCA-----CAACACTACTTTGGCCCTTCT 1890
QY 201 ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220
Db 1891 GGCATCTATTTCGCAAAATCTAATGACCTGCTTTACGTTATCGCAGAGCTTAAACCTT 1950
QY 221 LeuAsnSerTyrHisGlnGlnAla-----AspTyrLeuSerThrHisTyr 235
Db 1951 GTCAATGAGATTCAAAACGTTGATGGGTGTTTATTTTGTGTGTTTATATACCAACACAG 2010
QY 236 AlaGluGlnGlyAsnHisArgLeuPheGlu----- 245
Db 2011 GCTAGAAC-----TATCTCTTTGAAGATTGGAACCCCACTATATCAATTTTATG 2061
QY 246 -----AlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys 260
Db 2062 AATGCTGATGGTACGCTAGAGAGAGTTTGGCGAGGTCGTCTAGAACGACACATGCC 2121
QY 261 AspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluLeuLysGln 280
Db 2122 ATGAGTCGTGTATCTTGTGGTGGATTCGATTACTTATCTGACTGCTGAATTCAG----- 2175
```

```
QY 281 ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAsp 300
Db 2176 -----GTAGATGGTTTTCGTTTCACATGATGGGTGACCATGATGGCGCAGCTATTGAG 2229
QY 301 IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln 320
Db 2230 -----CAAGCCTTTAAGGCAGCCAAAGCCATTAA 2259
QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
Db 2259 ----- 2259
QY 341 TyrAsnThrProMetPheGlyAspSerTrpIleThr-----AspLysAsnPhe 356
Db 2260 CCAATATACCATTATGATTGGCAAGGCTGCGTACTTACCAAGGTGATGAGGGCAAAAA 2319
QY 357 ArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyr 376
Db 2320 GAAATTCGCGCAGATCAAGATTGGATGAAA-----GCAACCAATACGTCGGTGT 2370
QY 377 PheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsn 396
Db 2371 TTCTCTGATGATATCAGA-----AATACCCTCAAGTCAGGTTTTTCCAAAT 2415
QY 397 AlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsn-----AlaThrValMetVal 414
Db 2416 GAAGGCACAGCAGCCTTTTATTACTGTGGCGCAAAAAATCTAGAAGGTTTTTATCAAAACG 2475
QY 415 LeuLysAlaSerPro-----ProGly-----GluPheHis 424
Db 2476 ATCAAGCACAGCTGGTAACCTTTGAAGCAGATGCCCCAGGAGATGTAGTCAGATATT 2535
QY 425 AlaGlnProAspAsnGlyThrPhe---GluLeuPheIleLysGlyArgAsnPheThrPro 443
Db 2536 GCAGCCCATGACACACCTGACCTTACATGATCATTTGCCAAATCCATCAATAAGGATCCT 2595
QY 444 AspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyr 463
Db 2596 AAAGTG-----GCTGAAGAAGAGATT-----CAC 2619
QY 464 ArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLys 483
Db 2620 AAGCGTTATGCTAGCAAAATACCATGATTTTAACTGCTCAAGGACTGCTTTATCCAT 2679
QY 484 AlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnPro 503
Db 2680 TCTGGTCAG-----GAAATGACGAACCAACAGCAGCTTCTA-----AATCCC 2721
QY 504 SerTyr-----ProAsnLeuAspHisGlnArgSerVal 514
Db 2722 GACTACAAGACAAAGCGCTGATGACAAAGGTGCCAAAT-----AAGGCGACT 2769
QY 515 LeuPhe-----IleAsnLysLysTyrPheLeuValIleAspArgAlaIleGlyGlu 531
Db 2770 CTGATTGATGCTAGCGCAATACCTTACTTCAATCCAGATTCTTATGATTCGTCTGAT 2829
QY 532 AlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSer-----AsnProVal 549
Db 2830 GCGGTCAATCATTTT-----GACTGGGCAAGGCAACAGATTCCATAGCTCCCGGATT 2883
QY 550 PheAspLysThrLysAsnArgValTyrThr-----ThrTyrArgAspGlyAsn 565
Db 2884 AGCAACCAACAAAA-----GCCATATACAGGGAGTAATTCGTTGCTGCTCAACA 2937
QY 566 AsnLeuMetIleGlnSerLeuAsnAlaAsp----- 575
Db 2938 GATGCTTTTACAAAGCAACCAAGCTAGGTAGATCGGGATGTGACCTTGTATCCCCAA 2997
QY 576 -----ArgThrSerLeuAsnGluGluGlyLysValSerTyr-----ValTyrAsn 591
Db 2998 GCAGGACAGATGGTATTTCACAAAGGACCTCATCATGGGTACCACACATGGGCATCA 3057
QY 592 LysGluLeuLysArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGln 611
```


Db 94 GTATTATCTTTGCTATAGTCGTAATGAGTAGGTATTATTAACAATAGCAACAAGATCT 153
Qy 28 SerIleThrArg-----LysAspPheAspHisIleAsnLeuGluTyrSerGlyLeuGlu 45
Db 154 CATTACAAACAATTAATCAAAACGATGAACATAACT----- 192
Qy 46 LysValasnLysAlaValAlaAlaGlyAsn-----Tyr 56
Db 193 ---GTTAATGATCAGTACCAAGAGCCGAATACTAGATAGAAATGGCAAGTACTAGT 249
Qy 57 AspAspAlaAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluPro 76
Db 250 GATAATGCTTCAAGATGCTTATACATACACTAGAAACCGTAAACACATCAACAAGGAA 309
Qy 77 AspPheSerAsnAlaGluLysProAlaAsp---IleArgGlnProIleAspLysValThr 95
Db 310 ATGTTAAATACCTGTAAGAACTGACAGATTTAAATTAATGATGATACATGATTAATCT 369
Qy 96 ArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyr 115
Db 370 GAGAGA-----GATAAA----- 381
Qy 116 PheAspTyrGlyLysAspIleAsnTrpGlnMetTrpPro-----ValLysAspAsn 132
Db 382 -----AAGGATTTTGGATTCAATGATGATCCGTCATTTGCTAAAAAGTT-AAAT 428
Qy 133 GluValArgTrpGlnLeuHisArgValLysTrpTrpGln----- 145
Db 429 GAGAAAGA-----ACAATTAATGTTAGAGATGCGATTTTCAAGAGCAATTTGA 482
Qy 146 -----AlaMet 147
Db 483 TACCCAACTTACAGATAAAATAGGAAAAAAACAATTTAAAAACAGTTAACTAAAAAGATT 542
Qy 148 AlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla----- 160
Db 543 GCAAGTTTGTAGCAATTTATCGGGAATGAACGTCGGTCAACTCTAGATCTCAACAAT 602
Qy 161 -----ArgGlu----- 162
Db 603 TAAAAATGAGACGTAAAGGAGAGAGATATGACGCCGTATCACACACCTTTCTAAAT 662
Qy 163 ---TrpValTyrGlnTyrSer-AspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAs 181
Db 663 ACCTGGTGTAATACTACAATGATGGGATAGAAAAATACCCATACGGT-----GA 713
Qy 181 pAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProPr 201
Db 714 TACTTTAAGAGGCATATTTTGGA-----GATGTGCGACTTCGACTGGAAGGTATACCTAA 767
Qy 201 othrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLe 221
Db 768 AGAATTAAT-----GAACAATATT 788
Qy 221 uAsnSerTyrHisGlnGlnAlaAspTyrIleuSerThrHisTyrAlaGluGlnGlyAsnHi 241
Db 789 ATCAAAAGGTATTACGAAATGATCGGTGCGTAAATCTTATCTTGAATATCAATACGA 848
Qy 241 sArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAs 261
Db 849 AGATGTTCTTAAGACGACGAGAAACAATG-----AAATATACAC 890
Qy 261 pSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleLysLysGlnVa 281
Db 891 TGATAAATCTGGAAGAGTAATAAGTTTCAGAACTACTTAATCTCGGCTCAAGA----- 942
Qy 281 lTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspTl 301
Db 943 -----GGTCATGATTTACAATTAAT-----ATAGATAT 971
Qy 301 ePheLeuLysAlaTyrGlySerAlaLysArgValAsn-----LeuGluLysGluPhePr 319
Db 972 TGAATTACAG-----AAAAAGTAGATCTTTTATTAGAAAAACAATTTTC 1016

Qy 319 oGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuPr 339
Db 1017 TAAATACGTAGTCAAGGTGCTAAGATATGACAAATCGTGAATGGTTGTCACAAATCC 1076
Qy 339 oAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAl 359
Db 1077 TAAAAATGGAGACATCTCGCTATTGCAGGAAGCAAAATGTATAGCAAGGTAAACTCAA 1136
Qy 359 aGlnPhe-----AlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLys---Ty 376
Db 1137 AGATTATGATATCGCAACTTTTACAGCTCAATACACAGTAGGTCTTCTCAGTAAAGAGG 1196
Qy 376 rPheAlaThrAspGlyLysGlnGlyLysAlaProAsn-----PheLeuSerLys 392
Db 1197 AACATTATTAGCTGGATACCAAAATTAAGCTATTATGTGTGGAGAAACTATGTAGATGA 1256
Qy 392 sAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAla----- 410
Db 1257 GCCATTAAAAATCCAAAGGTGGTTTAACCTAAGCGTCTTATTATTAATAAAATGGTCATGT 1316
Qy 411 -----ThrVa 412
Db 1317 ATCTATCGATGATAAACAAGCACTTATGCAITTCATCAAAACGTATACATGTTTAAACCGC 1376
Qy 412 lMetValLeuLysAlaSerProProGlyGluPheHisAlaGlnProAspAsnGlyThrPh 432
Db 1377 ACTTAAATTAGCAGGTGACCCCTATCTTACGTATGTCATTACCTTAATAT----- 1428
Qy 432 eGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGl 452
Db 1429 -----ATACGACATGCTGT----- 1443
Qy 452 yAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLe 472
Db 1444 -----CGTAAATTCGTAAGGGTTAAATCAAGTAGTCTTGGGCTTAAAAAC 1490
Qy 472 uThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGl 492
Db 1491 AGGTATTGACTTACCGACGAAACCCGACGCAATAGAACCTTAACTAAATCTCTGG 1550
Qy 492 yAsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnAr 512
Db 1551 TAACTATTTAGACTTAGCTATTGGACAATACACACATATACACCACTTTCAGTTGCCCA 1610
Qy 512 gSerValLeuPheIleAsnLysLysTyrPheLeuVal-----IleAspArgAlaTl 529
Db 1611 ATACGTATCAACTATTGCTAATGATGGCTATAGAATTCAAACACATATTGGATTGTCTAT 1670
Qy 529 eGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLysGluAspSerAsnProVa 549
Db 1671 TTATGAATCTACTAAT-----AAAGATGAACCTGGTCCATT 1706
Qy 549 lPheAspLysThrLysAsnArgValTyrThrTyrArgAspGlyAsnAsnLeuMetIl 569
Db 1707 AAAACGCTAAAAATTAAGGGTAAATGTTTAAATTAAGGTAAATAACTCAATAGCAGAAATTA 1766
Qy 569 eGlnSerLeuAsnAlaAspArgThrSerLeuAsnGluGluGlyLysValSerTyrV 589
Db 1767 AGAAGTTCAGAAAGGCTTCAAAATGCTTTCATGAAGCAAGGTACAGGTATGTCTAG 1826
Qy 589 al-----TyrAsnLysGluLeuLysArgProAlaPheValPheGluLysP 604
Db 1827 TTTTAGAATACTGTAGTACCTTCAGCTGGTGAACAGGAAGTCTGTAAGTTTTCAGA 1886
Qy 604 roLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrPro 619
Db 1887 CGGAGAACCTAGAGTTAACTCAACAT---ATATCGTATTATGCACCCG 1930

RESULT 12

AAN40009

ID AAN40009 standard; DNA; 3994 BP.

XX

AC AAN40009;
 XX 30-NOV-1991 (first entry)
 DE Saccharomyces cerevisiae invertase gene.
 XX Signal sequence; invertase; vector; leukocyte interferon.
 XX Saccharomyces cerevisiae.
 XX Key Location/Qualifiers
 FT 1896..3494
 FT /*tag= a
 XX
 XX EP127304-A.
 XX
 XX 05-DEC-1984.
 XX
 XX 24-APR-1984; 84EP-0302722.
 XX
 XX 25-APR-1983; 83US-0488337.
 XX (GETH) GENENTECH INC.
 XX
 XX Chang CN, Matteucci MD, Hitzeman RA;
 XX WPI; 1984-301996/49.
 XX P-PSDB; AAP40020.
 XX
 XX Yeast expression vehicle - contains DNA sequence of yeast promoter
 PT linked to DNA sequence encoding homologous protein.
 XX
 XX Disclosure; Fig. 13; 53pp; English.
 XX
 XX The signal sequence of *S. cerevisiae* invertase may be used in
 CC a new yeast expression vector for the expression of, esp. human
 CC leukocyte interferon-alpha. The expression vector comprises a yeast
 CC promoter operably linked to the signal sequence. This is attached
 CC to the coding sequence of the protein. Yeasts transformed with this
 CC vector produce and secrete protein normally heterologous to them.
 CC the protein can be recovered from the medium free of presquences
 CC etc. Human leukocyte interferon can be used as an antiviral or
 CC antitumour agent. The promoter is esp. the invertase promoter or
 CC the yeast 3-phosphoglycerate kinase promoter.
 XX
 XX Sequence 3994 BP; 1201 A; 773 C; 767 G; 1253 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 0.0573 Length: 3994
 Score: 127.00 Matches: 118
 Percent Similarity: 32.34% Conservative: 78
 Best Local Similarity: 19.47% Mismatches: 185
 Query Match: 3.63% Indels: 225
 DB: 5 Gaps: 33
 US-09-802-285A-2 (1-659) x AAN40009 (1-3994)
 QY 94 ValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyr 113
 DB 1956 ATGACAAACGAACTAGCATGACCTTTGGTCCAC---TTACACCCCAACAAAGGCGCTGG 2012
 QY 114 -----GlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrp--- 127
 DB 2013 ATGAATGACCCAAATGGTGTGGTACGATGAAGAAGATGCCAATGGCATCTGTACTTT 2072
 QY 128 ProValLysAspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMet 147
 DB 2073 CAATACAAACCCAAATGACACCGCATGG-----GGTACGCCATGTTTGGGGC----- 2120
 QY 148 AlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpValTyrGlnTyr 167
 DB 2121 -----CATGCTACTTCCGATGAT-----TTG 2141

QY 168 SerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTrp 187
 DB 2142 ACTAATTTGG---GAAGATCAACCCATGTCTATCGCTCCCAAGCGTAAC----- 2186
 QY 188 ArgProLeuGluValSerAspArgValGlnSerLeuProProThrPheSerLeuPheVal 207
 DB 2186 ----- 2186
 QY 208 AsnSerProAlaPheThrProAlaPheLeuMetGluPhe----- 220
 DB 2187 GATTACAGGTGCTTTCTCTGGCTCCATGGTGGTTGATTACAAACAACAGAGTGGGTTTTC 2246
 QY 221 -----LeuAsnSerTyrHis 225
 DB 2247 AATGATACTATTGATCCAGACAAAGATGCGTTGGGACTTTATTAACACTCTCTGAA 2306
 QY 226 GlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGlu 245
 DB 2307 AGTGAAGAGCAATACATTAGC-----TATCTCTTGTGGTGTGTACACTTTTACTGAA 2360
 QY 246 AlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAsp----- 261
 DB 2361 TACCAAAAGAACCCCTGTTTAGCTGCCAACTCCACTCAATTCAAGATCCAAAGGTGTTTC 2420
 QY 262 -----SerProArgTrpArgGlnThrGlyLeuSerValLeuAsnThrGluIle 277
 DB 2421 TGGTATGAACCTTCTCAAAATGGATTATGACG-----GCTGCCAAATCAACAAGACTAC 2474
 QY 278 LysLysGlnValTyrAlaAsp----- 284
 DB 2475 AAAATGAAATTTACTCTCTGATGACTTGAAGTCTCTGGAAGTAGAATCGATTGCTGCC 2534
 QY 285 -----GlyMetGlnPheGluLeu----- 290
 DB 2535 AATGAAGGTTTCTTAGGCTACCAATACGAATGTCCAGGTTTGATGCAAGTCCCAACTGAG 2594
 QY 291 -----SerProIleTyrHisValAlaAlaIleAspIle----- 301
 DB 2595 CAAGATCCTTCCAAATCTTATTGGGTCAATGTTTATTCTATCAACCCAGGTGCACCTGCT 2654
 QY 302 -----PheLeuLysAlaTyr 306
 DB 2655 GCGGGTTCCTTCAACCAATATTTTGTGGATCCTTCAATGGTACTCATTTGAAGGTTT 2714
 QY 307 GlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyr-----ValGln 324
 DB 2715 GACAATCAATCTAGAGTG-----GTAGATTTTGGTAAGGACTACTATGCTTGCAA 2765
 QY 325 ThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnThrPro 344
 DB 2766 ACTTCTTCAAC-----ACTGACCCCAACCTTACCGTTTCAGCA 2801
 QY 345 MetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPheAla----- 362
 DB 2802 TTA---GGTATTGCTCGGCTTCA-----AACTGGGAGTACAGTGCCTTTGTCCCACT 2852
 QY 363 -----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLys 375
 DB 2853 AACCCATGGAGATCATCCATGTCTTTGGTCCCAAGTTTCTTTGAAC-----ACTGAA 2906
 QY 376 TyrPheAla-----ThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLys 392
 DB 2907 TATCAAGCTAATCCAGAGACTGAATTTGATCAATTTGAAGCGCAACCAATATTGTAAC--- 2963
 QY 393 AlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTyrAspLysAsnAlaThrVal 412
 DB 2964 ---ATTAGTAATGCTGGT-----CCCTGGTCTCGTTTGGTACTAAC 3002
 QY 413 MetValLeuLysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPhe 432
 DB 3003 ACAACTCTCTAAGGCCCAATTTCTTACAATGTGCGATTGAGCAACTCGACTGGTACCTTA 3062
 QY 433 GluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGly 452

Db	3063	GAGTTTGAGTGG- ::: 	::: GTTTACGCTGTAAACACC	3092
QY	453	AspGluAlaIleMetLys- ::: 	--LeuArgAsnTrpTyrAsgGlnThrArg ::: 	467
Db	3093	ACACAACCAATCCAAATCGGCTTTGCCGACTTATCATTGGTTCAAGGGTTTAGAA ::: 	3152	
QY	468	- - - - - 	--IleHisSerThrLeuThrLeuAsp ::: 	475
Db	3153	GATCCTGAAGAATATTTGAGAATGGGTTTTGAAGTCAGTCTTCTCTCTCTTTTGGAC 	3212	
QY	476	AsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGly- ::: 	492	
Db	3213	CGTGGTAACCTAAGGTCAAGTTTGTCAAGGAGAACCCATATTTTCACAAAAGAGATGTCT ::: 	3272	
QY	493	- - - AsnAsnLeuaspValLeuThrTyrThrAsnProSerTyrProasn- ::: 	507	
Db	3273	GTCAACAAACCAACCATTCAGTCTGAGAACGCCTTAAGTTACTATAAAGTGTACGGCCTA 	3332	
QY	508	LeuAspHisGlnArgSerValLeuPheIleAsn- 	LysLys 520	
Db	3333	CTGGATCAAACATCTTTGGAATTTGACTTCAACGATGGAGATGTGGTTTCTACAAAATACC 	3392	
QY	521	TyrPheLeuValIleaspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrp ::: 	540	
Db	3393	TACTTCATGACCACCGGTAAACGCTCTAGGATCTGTGAACATGACCACCTGGTGTC- ::: 	3446	
QY	541	GlnLeuLysGluaspSerAsnProValPheAspLys-ThrLys- ::: 	AsnAr 556	
Db	3447	- - - - - GATAATTGTCTCATTTGACAAGTTCCAAGTAAGGGAGTAAATAAG 	3494	
QY	556	gValTyrThrTyr 561 		
Db	3495	AGGTTATAAACATTAT 3510 		
<hr/>				
RESULT 13				
AAQ11883				
ID	AAQ11883 standard; DNA; 3994 BP.			
XX	AAQ11883;			
XX				
DT	24-JUL-1991 (first entry)			
XX				
DE	Invertase gene.			
XX				
KW	Invertase; yeast; promoter; expression vehicle; interferon; ss.			
XX				
OS	Saccharomyces cerevisiae.			
Key	Location/Qualifiers			
CDS	1896..3491			
FT	/*tag= a			
FT	/product= invertase			
XX				
PN	US5010003-A.			
XX				
PD	23-APR-1991.			
XX				
Pf	20-JUN-1990; 90US-0541186.			
XX				
PR	25-APR-1983; 83US-0488387.			
PR	20-JUN-1990; 90US-0541184.			
XX				
PA	(GETH) GENENTECH INC.			
XX				
PI	Chang CN, Hitzeman RA, Matteucci MD;			
XX				
DR	WPI; 1991-148222/20.			
DR	P-PSDB; AAR12145.			
XX				
Pr	Yeast expression vehicle - contg. homologous yeast signal peptide			
PT	for processing and export of mature heterologous protein from			
PT	yeast cells			


```

QY 285 -----GlyMetGlnPheGluLeu----- 290
Db 2535 AATGAAGTTTCTTAGCTACCAATACGAATGCCAGGTGATGAAGTCCCAACTGAG 2594
QY 291 -----SerProIleTyrHisValAlaAlaIleAspIle----- 301
Db 2595 CAAGATCCTTCCAAATCTTATGGTCTATGTTTATTTCTATCAACCCAGGTGCACCTGCT 2654
QY 302 -----PheLeuLysAlaTyr 306
Db 2655 GCGGTTTCCTTCAACCAATATTTTGTGGATCCCTCAATGGTACTCATTTTGAAGCGTTT 2714
QY 307 GlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyr-----ValGln 324
Db 2715 GACATCAATCTAGAGTG-----GTAGATTTTGGTAAGACTACTATGCTTGCATTCGAA 2765
QY 325 ThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnThrPro 344
Db 2766 ACTTCTTCAAC-----ACTGACCCCAACTACGGTTCAGCA 2801
QY 345 MetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPheAla----- 362
Db 2802 TTA---GGTATGCTGGCTTCA-----AACTGGGAGTACAGTGCCTTTGTCCCAACT 2852
QY 363 -----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLys 375
Db 2853 AACCCATGGAGATCATCATCTTTGGTCCGCAAGTTTCTTTGAAC-----ACTGAA 2906
QY 376 TyrPheAla-----ThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLys 392
Db 2907 TATCAAGCTAATCCAGAGACTGAATGATCAATTTGAAGCCGCAACCAATATTGAAC--- 2963
QY 393 AlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrVal 412
Db 2964 ---ATTAGTAATGCTGGT-----CCCTGGTCTCGTTTGTGTACTTAAC 3002
QY 413 MetValLeuLysAlaSerProProGlyGluPheHisAlaGlnProAsnPheThrPhe 432
Db 3003 ACAACTCTAACTAAGGCAATCTTACATGTGATGTGAGCAACTGACTGTGACCTTA 3062
QY 433 GluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGly 452
Db 3063 GAGTTTGAGTTG-----GTTTACGCTGTATACACC 3092
QY 453 AspGluAlaIleMetLys-----LeuArgAsnTrpTyrArgGlnThrArg 467
Db 3093 ACACAAACCAATATCCAAATCCGCTTTTGGCGACTTATCACTTTGGTTCAGGGGTTAGAA 3152
QY 468 -----IleHisSerThrLeuThrLeuAsp 475
Db 3153 GATCCTGAGAAATATTGAGAATGGGTTTGAAGTCAGTCTCTCTCTCTTTTGGAC 3212
QY 476 AsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGly----- 492
Db 3213 CGTGGTAACTTAAGTCAAGTTTGTCAAGGAGAACCCATATTTCACAAACAGATGTCT 3272
QY 493 ---AsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrTrpProAsn----- 507
Db 3273 GTCACACCAACCACTTCAAGTCTGAGAAGCACTAAGTTACTATAAAGTGTACGGCCTA 3332
QY 508 LeuAspHisGlnArgSerValLeuPheIleAsn-----LysLys 520
Db 3333 CTGGATCAAAACACTCTTGGNAATGTACTTCAACGATGGAGATGTGGTTCTTACAAATACC 3392
QY 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrp 540
Db 3393 TACTTCATGACCCCGCTAAGCTCTAGGATCTGTGAACATGACCACTGGTGTCT----- 3446
QY 541 GlnLeuLysGluAspSerAsnProValPheAspLys-ThrLys-----AsnAr 556
Db 3447 -----GATAATTGTTTCTACATTGACAAAGTTCACAGTAAGGGAAGTAAATAG 3494

```

```

QY 556 gValTyrThrThrTyr 561
Db 3495 AGGTATAAAACTTAT 3510
RESULT 14
AAQ41744
ID AAQ41744 standard; DNA; 3414 BP.
XX AAQ41744;
AC AAQ41744;
XX
XX 14-SEP-1993 (first entry)
XX
XX DNA encoding B.t. toxin HD511.
XX Coleopteran pests; Bacillus thuringiensis; insecticide; transgenic;
XX plant; pesticide; Colorado potato beetle; ss.
XX Bacillus thuringiensis strain HD511.
XX WO9308693-A.
XX 13-MAY-1993.
XX 06-NOV-1992; 92WO-US09510.
XX 06-NOV-1991; 91US-0788638.
XX (MYCO ) MYCOGEN CORP.
XX Fu JM, Payne JM;
XX
XX WPI; 1993-167285/20.
XX P-PSDB; AAR37213.
XX
XX Control of coleopteran insect pests using Bacillus thuringiensis
XX - using strains Bt HD 511, HD 867 or HD 1011 or their toxins or
XX nucleotide sequences
XX
XX Claim 13; Page 13-15; 29pp; English.
XX
XX The DNA sequence encoding Bacillus thuringiensis strain HD 511 toxin
XX may be used in an insecticide/pesticide for treatment of Coleopteran
XX pests, esp. the Colorado potato beetle. Strain HD511 and recombinant
XX cells contg. the gene encoding the toxin may be treated by chemical or
XX physical means to prolong the pesticidal activity of the cells, the
XX treated cells acting as a protective coating for the pesticidal toxin,
XX which becomes available upon ingestion by a target insect. The compsn.
XX may be applied to the environment of the coleoptera, e.g. plants, soil
XX or water, by spraying, dusting, etc.
XX See also AAQ41745.
XX
SQ Sequence 3414 BP; 1215 A; 534 C; 658 G; 1007 T; 0 other;
Alignment Scores:
Pred. No.: 0.101 Length: 3414
Score: 123.50 Matches: 141
Percent Similarity: 35.92% Conservative: 95
Best Local Similarity: 21.46% Mismatches: 199
Query Match: 3.53% Indels: 224
DB: 14 Gaps: 39
US-09-802-285A-2 (1-659) x AAQ41744 (1-3414)
QY 48 AsnLysAlaValAla-----AlaGlyAsnTyrAspAlaAlaLysAlaLeu 63
Db 361 AATAAGCACTTGCAGAAATTAGAGGATTAGGAATAAC----- 399
QY 64 LeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsnAlaGluLys 83
Db 400 TTAACAATATATCAACAG-----GCACCTTGAAGATTGGCTGAACAATCTGTATGAT 450
QY 84 ProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAspLysAlaLeu 103
Db

```

```
Db 451 CCAGCACTATAACACGAGTGATAGATCGTTT-----CGTATTATTAGAT---GCTTTA 501
QY 104 ValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAsp----- 117
Db 502 TTGGAATCATATATCCCGTCATTATAGGGTGTCTGGATATGAATAATCACTAACAAGT 561
QY 118 TyrGlyLysAspIleAsnTrpGlnMet----- 126
Db 562 TACGCACAAAGCGGCAACCTTCATCTAGCTTTATTAAAGAGATTCCTACTCTTTATGGAGAT 621
QY 127 ---TrpProValLysAspAsnGluVal----- 134
Db 622 AAATGGGATTCACTCAGAACACATTTAGGAAAATTATAATCGTCAAAAGAAACATATC 681
QY 135 ---ArgTrpGlnLeuHisArgValLysTrpTrpGlnAla---MetAlaLeuValTyrHis 152
Db 682 TCTGAATATCTAACCAATTCGCTTAAGTGGTATAATAGTGGCTTAGCGAGATGAACCGT 741
QY 153 AlaThrGlyAspGlu-----LysTyrAlaArgGluTrpValTyrGlnTyr 167
Db 742 TCCACTTATGAACATGGATAAATTATATATCGTTTCGTAGAGAAATGATATTAATGGTA 801
QY 168 SerAspTrpAlaArgLysAsnProLeu-----GlyLeu 178
Db 802 TTAGATATTGCTGTATTCTCTATTATGACCCCTCGAATGTAATTCATATGGAACAACT 861
QY 179 SerGlnAspAsnAspLysPheValTrpArg-ProLeuGluValSerAspArgValGlnSe 198
Db 862 ACGCAGTT-AACGAGAGAAGTGTATACCGCATCAATTAGCTTGCTCAATTAGCAATCCAGA 920
QY 198 rLeuProTrpPheSerLeuPheValAsnSerProAlaPheTrpAlaPheLeuMe 218
Db 921 TATAGTCCAGCTTTTCTCAGATGGAAATACTGCGTTTAGAACACACACCTTGTT-- 978
QY 218 tGluPheLeuAsnSerTyrHisGlnAlaAlaAspTyrLeuSerThrHisTyrAlaGlu 238
Db 979 -----GATTATTATAGATGAGCTTTATATATATATAC 1007
QY 238 nGlyAsnHisArgLeuPhe-----GluAlaGlnArgAsnLeuPhe-----Al 252
Db 1008 ATCAAAATAAAGCATTTTTCATGAGATTCAACGACCTATTATTTATGGTGTGTACA 1067
QY 252 aGlyValSerPheProGluPheLysAspSerProArgTrpArgGlnThrGlyLe-SerV 272
Db 1068 TAAGGTACTTTTAAATAATCGGAGCAATCCAATTATAT--ACAACAGGCATATATG 1124
QY 272 aLeuAsnThrGluLe-----LysLysGlnValTyrAlaAspGlyMet---GlnPheG 289
Db 1125 TAAACAGTGGATATATTTCATCAGGAGCATATTCATTTAGAGGGAATGATATCTATAG 1184
QY 289 luleuSerProIleTyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerA 309
Db 1185 ACATTAGCAGCTCCATCATGTTGTAGTTTATCCGTATA---CTCAGAAATATGTCGTCA 1241
QY 309 lalys----- 310
Db 1242 GCAAGTTGAGTTTACGGTGTAAAGGCGCATGPACATTATAGAGGAGATAACAATAATGA 1301
QY 311 ---ArgValAsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnM 329
Db 1302 TCTGAGTATGATTCTTATTGATCAATTACCCCGAGGAGAACCAACCAACAGAAAATA 1361
QY 329 etileMetAlaLeu-----IleSerIleSerLeuProAspTyrAsnThr- 343
Db 1362 CACTCATCGA-TTATGTCATGTACAGCTATATCTAAATCAACTCCGGATTATGATAATG 1420
QY 344 ---ProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnP 361
Db 1421 CTACTATCCGATCTTT-----TCTTGG-----ACGCATAGAAGTGGCGAGT 1462
QY 361 heAlaSerTrpAlaArgValPheProAlaAsn-----GlnAlaIleLysTyrP 377
Db 1463 AT-----TACAATAGATCTATCCAAACAAATAATCAAAAAATTCAGCTGTAAAAATGT 1516
```

```
QY 377 heAlaThrAspGlyLysGlnGly-----LysAlaProAsnPheLeuSerLysAlaLeuS 395
Db 1517 ATAAACTAGATCAATCTACTACAGTTGTCAAAAGGCGCTGGATTACAGGTGAGATTAG 1576
QY 395 erAsnAlaGlyPheTyrThrPheArgSerGlyTyr-----AspLysAsnAlaThrValM 413
Db 1577 TTAAGAGAGGG-----AGTAATGGTTATATAGGAGATATAAAGGCTACCGTA- 1623
QY 413 etValLeuLysAlaSerProProGlyGlu-----PheHisAlaGlnP 427
Db 1624 -----AATCAACCACTTCTCAAAATATCGTGTAGAGTTTCATACGCCACTA 1672
QY 427 toAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyV 447
Db 1673 GTGTTTCTGGACTATTCAACGTTTAT----- 1701
QY 447 alPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrA 467
Db 1701 ----- 1701
QY 467 rgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnA 487
Db 1702 -----AATGATGAATAGCGCTTCAAAAAAATTTTCAAA 1735
QY 487 snLysTrpGluThr---GlyAsnAsnLeuAspValIleThrTyr-----LeuA 509
Db 1736 GTACTGTAGAACAAATAGGTGAAGGAAAGAT---TTAACTATGTTTCATTTGGATATA 1792
QY 501 -----ThrAsnProSerTyrProAsn----- 500
Db 1793 TAGAATATTCTACGACCATTCATTTCCGAATGAGCATCCAAAAATCACTTCTCATTTAA 1852
QY 509 spHis-----GlnArgSerValPheIleAsnLysTyrPheLeuValIleAspA 527
Db 1853 ACCATTGAGTAAACAATTCACCATTATATGTAGATTCATCACTGTAATTTATCCCTGTAGAT- 1911
QY 527 rgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerA 547
Db 1912 -----GTAAATTATGATGAAAAAGAAAAA--- 1935
QY 547 snProValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGlyAsnAsnL 567
Db 1936 -----CTAGAAAAAGCACAGAAAGCGGTGAATACCTTGTTCACAGAGGGAAGAATG 1987
QY 567 euMet-----IleGlnSerLeuAsnAlaAspArgThrSerLeu 579
Db 1988 CACTCCAAAAATACGTGACAGATTATAAGTGGACCGGTTTCAATT 2034

RESULT 15
AAV09317
ID AAV09317 standard; cDNA; 1542 BP.
XX AAV09317;
DT 13-MAY-1998 (first entry)
DE Unsecreted yeast invertase gene (SUC2) used as a selection marker.
XX Secreted protein; unsecreted invertase; yeast; selection marker;
KW isolation; ss.
OS Saccharomyces cerevisiae.
XX
XX US5712116-A.
DN
PD 27-JAN-1998.
XX
XX 24-APR-1996; 96US-0639052.
XX
PR 07-APR-1993; 93US-0045267.
PR 24-OCT-1994; 94US-0328962.
PR 24-APR-1996; 96US-0639052.
```


QY 508 LeuaspHisGlnArgSerValLeuPheIleasn-----LysLys 520
 Db 1381 CTGGATCAAAACATCTTGGATGTGACTTCAACGATGGAGATGGTTTCTACAAATACC 1440
 QY 521 TyrPheLeuValIleaspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyVal 538
 Db 1441 TACTTCATGACCACCGGTAAACGCTTAGGATCTGTCAACATGACCACCTGGTCTC 1494

Search completed: August 4, 2003, 11:30:38
 Job time : 4005 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2003, 07:43:34 ; Search time 4666 Seconds
(without alignments)
4110.319 Million cell updates/sec

Title: US-09-802-285A-2

Perfect score: 3494

Sequence: 1 MTKIFKRIIVFAVIALSSG.....KGLNLTTLTKGQQLVLVP 659

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09802285/runat_28072003.190510.9705/app.query.fasta_1.839
-DB=GenEmbl -QFMT=fastap -SURFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09802285@cgn.1.1.3052 @runat_28072003.190510.9705 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3494	100.0	1980	1	CHU27586	U27586 Cytophaga h
2	3494	100.0	1980	6	I71365	I71365 Sequence 3
C 3	235.5	6.7	50327	1	AE014145	AE014145 Streptoco
C 4	234.5	6.7	12195	1	AE006517	AE006517 Streptoco
C 5	233.5	6.7	12187	1	AE010002	AE010002 Streptoco
C 6	198	5.7	175936	2	SPNEU1908	AL449930 Streptoco
C 7	189	5.4	11333	1	AE008410	AE008410 Streptoco
C 8	186.5	5.3	34496	1	SCF62	AL121855 Streptomy
C 9	177.5	5.1	22913	1	AF498407	AF498407 Pseudomon
C 10	176	5.0	3763	6	BD003866	BD003866 Polynucle
C 11	176	5.0	14385	1	AE007344	AE007344 Streptoco
C 12	176	5.0	19295	1	AF498414	AF498414 Pseudomon
C 13	170.5	4.9	16720	1	AF498401	AF498401 Pseudomon
C 14	170.5	4.9	16720	1	AF498411	AF498411 Pseudomon
C 15	160.5	4.6	205050	1	AL646082	AL646082 Ralstonia
C 16	154	4.4	2882	1	AB003330	AB003330 Pseudomon
C 17	145.5	4.2	10566	1	AE010107	AE010107 Streptoco
C 18	140.5	4.0	50354	1	AE014169	AE014169 Streptoco
C 19	139	4.0	2295	6	AX415476	AX415476 Sequence
C 20	139	4.0	23292	1	SCE134	AL049661 Streptomy
C 21	139	4.0	258650	1	AL596171	AL596171 Listeria
C 22	139	4.0	349980	6	AX417046	AX417046 Sequence
C 23	139	4.0	349980	6	AX417047	AX417047 Sequence
C 24	137.5	3.9	10586	1	AE006620	AE006620 Streptoco
C 25	137.5	3.9	26953	1	ALW243431	AJ243431 Acinetoba
C 26	134	3.8	321250	1	MPULM02	AL445564 Mycoplasma
C 27	132.5	3.8	67970	3	PFMAL1P3	AL031746 Plasmodiu
C 28	129.5	3.7	39305	8	SPU33009	U33009 Schizosacch
C 29	129.5	3.7	41799	8	SPFC1683	AL355920 Schizosac
C 30	129.5	3.7	85837	8	SPU33010	U33010 Schizosacch
C 31	129	3.7	9130	12	AF138275	AF138275 Synthetic
C 32	129	3.7	9193	12	AF138273	AF138273 Synthetic
C 33	129	3.7	9194	12	AF138274	AF138274 Synthetic
C 34	128	3.7	13280	1	AE013673	AE013673 Yersinia
C 35	128	3.7	148753	2	AC122951	AC122951 Rattus no
C 36	128	3.7	216050	1	AJ414157	AJ414157 Yersinia
C 37	127	3.6	1801	8	AV128845	AV128845 Arabidops
C 38	127	3.6	1993	8	AY081353	AY081353 Arabidops
C 39	127	3.6	2873	8	SCINVE	VO1311 Yeast gene
C 40	127	3.6	3994	6	E00284	E00284 DNA coding
C 41	126	3.6	3749	1	AF305610	AF305610 Borrelia
C 42	125	3.6	3072	1	AF367012	AF367012 Streptoco
C 43	124	3.5	4410	1	LEUORF15X	L28806 Leuconostoc
C 44	124	3.5	4410	1	LOP10RFG	Z29976 L.oenos pla
C 45	123.5	3.5	3414	1	BTU04367	U04367 Bacillus th

ALIGNMENTS

RESULT 1

```

CHU27586
LOCUS      1980 bp      DNA      linear      BCT 05-NOV-1996
DEFINITION Cytophaga heparina heparinase III (HepC) gene, complete cds.
ACCESSION  U27586
VERSION     U27586.1
KEYWORDS   GI:924924
SOURCE     pedobacter heparinus.
ORGANISM   pedobacter heparinus
            Bacteria; Bacteroidetes; Sphingobacteriales; Sphingobacteriaceae; Pedobacter.
            1 (bases 1 to 1980)
            Su.H., Blain,F., Musil,R.A., Zimmermann,J.J., Gu,K. and
            Bennett,D.C.
            Isolation and expression in Escherichia coli of hepB and hepC,
            genes coding for the glycosaminoglycan-degrading enzymes heparinase
            II and heparinase III, respectively, from Flavobacterium heparinum
            Appl. Environ. Microbiol. 62 (8), 2723-2734 (1996)
            96316388
            8702264
            2 (bases 1 to 1980)
            Tkalec,A.L.
            Direct Submission
            Submitted (24-MAY-1995) Ana L. Tkalec, IBEX Technologies, 5485
            Pare, Montreal, Quebec, H4P 1P7, Canada
            Location/Qualifiers
            1..1980
            /organism="pedobacter heparinus"
            /db_xref="taxon:984"
            1..1980
            /gene="HepC"
            1..1980
            /gene="HepC"
            /EC_number="4.2.2.8"
            /protein_id="AA18278.1"
            /transl_start=1
            /product="heparinase III protein"
            /protein_id="AA18278.1"
            /db_xref="GI:924925"
            /translation="MTTKIFKRIIVFAVIALSSGNILAQSSITKDFDHINLEYSGL
            EKNKVAAGNYDDAKALLAYREKSKARPEDFNSAEKPADIRPIDKVTRERADKA
            LVHQPHKGYGFDYGDNDINQWMPVNDNEVRQHLRVKWWQAWALVYHATGDEKYA
            REWYQSDWARKNPLGSDQNDKFMVRPLEYSDRVLSLPTFLSFLVNSPAPTAPFLM
            EFLNSYHQOAYLSTHYAEQNHRLFEAQRNLFAGVSPFEKDSPRWQGTGIVLNTF
            IKQVADQMDFELSPIVHVAADIDFLKAYGSAKRNVLEKPEPQSYVQTVENMIMALI
            SISLPDNTPWGDSMTDKNPRMAQFASWARVFPANOAIKYFATDGGKAPNPLSK
            ALSNAGFTFSGWDKNATVMVLKASPGCFHAQPDNGTEFLFKGRNPTDAGVEVY
            SGDEAIMKLRWYRQTRIHTSLTLDNQNVITKARQNKWETGNLDLVITYTNPSPNLI
            DHQSVLPFKKKYFVUIDRAIGATGNLGVHQLKEDSNPVDKTKNRVYTYTYRDGNN
            LMTSLNADRTSLNEEGKVSXYNKEIKRPAPFVFKPKKNAGTQNFVSI VYPYDQK
            APEISIRENKGNDPEKKNLTTLTINGKQLVLP"
            sig_peptide 1..72
            /gene="HepC"
BASE COUNT 630 a 432 c 438 g 480 t
ORIGIN
Alignment Scores:
Pred. No.:      1,85e-280      Length:      1980
Score:          3494.00      Matches:      659
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              1          Gaps:      0
US-09-802-285A-2 (1-659) x CHU27586 (1-1980)
QY      1 MetThrThyLysIlePheLysArgIleIleValPheAlaValIleAlaLeuSerSerGly 20
DB      1 ATGACTACGAAATTTTAAAGGATCATTTGCTGTAATTGCTCCCTATCGTCGGGA 60
QY      21 AsnIleLeuAlaGlnSerSerIleThrArgLysAspPheAspHisIleAsnLeuGlu 40
DB      61 AATATACTTGCAAAAGCTCTTCCATACAGGAAAGATTTTGACCACATCAACCTTGAG 120
41 TyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAspAlaAla 60
121 TATTCCGGACTGGAAAAGGTTATAAAGCAGTTGCTCGCGCAACTATGACGATCGCGCC 180
61 LysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsn 80
181 AAAGCATTTACTGGCATACTACAGGAAAAGTAGTAGGCCAGGGAACCTGATTTCAGTAAT 240
81 AlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAsp 100
241 GCAGAAAGCCTGCCGATATACGCCAGCCCATAGATAGGTACGCGTGAATGSCCGAC 300
101 LysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLys 120
301 AAGGCTTTGGTCCACCATTTTCAACCGCACAAAGGCTACGCGCTATTTTGAATTATGGTAAA 360
121 AspIleAsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArg 140
361 GACATCAACTGGCAGATGTGGCGGTAAAAGACAAATGAAGTACGCTGGCAGTTGCACCGT 420
141 ValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla 160
421 GTAAATGTCGCGAGCTATGCGCTGTTTATCAGCTACGGCGCATGAAAATATATGCA 480
161 ArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGln 180
481 AGAGAAATGGGTATATCAGTACAGCGATTGGGCGCAGAAAAACCCCATTTGGCGCTGTCGCAG 540
181 AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro 200
541 GATAATGATAAATTTGTGTGGCGCCCTTGAAGTGTGGACAGGGTACAAAGTCTTCCC 600
201 ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220
601 CCAACCTTCAGCTTATTTGTATAACTCCGCGAGCTTTACCCAGCGCTTTTAAATGGAATTT 660
221 LeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsn 240
661 TTAAACAGTTTACCACCAACAGCGCGCATTTATTTATCTACGCATTATTCGCGAACAGGGAAC 720
241 HisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys 260
721 CACCGTTTATTGGAAGCCCAACGCAACTGTGTTCAGGGGTATCTTTCCCTGAAATTTAAA 780
261 AspSerProArgTyrArgGlnThrGlyIleSerValLeuAsnThrGluIleLysLysGln 280
781 GATTCACCAAGATGAGGCAACCCGCATATCGGTGCTGAACACCGAGATCAAAAAACAG 840
281 ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAsp 300
841 GTTTATGCCGATGGGATGCAGTTTGAACTTTTCAACCAATTTTACCATGTAGTGCCTATCGAT 900
301 IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln 320
901 ATCTTTCTTAAAGGCGCTATGGTTCGCAAAACAGATTAACTTTGAAAGAAATTTCCGCA 960
321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
961 TCTTATGTACAAACTGTAGAAAATATGATTATGGCGCTGATCAGTATTTTCACTGCCAGAT 1020
341 TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGln 360
1021 TATAACACCCCTATGTTTGGAGATTTCATGGATTACAGATATAAAATTTTTCAGGATGCACAG 1080
361 PheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAsp 380
1081 TTTGCGAGCTGGCGCGGTTTTCGCGCAACACGAGCCATATAATATTTTGTCTACAGAT 1140
381 GlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyr 400
1141 GCAAAACAGGTAAAGCGCGCTAACTTTTATCCAAAGCATTCAGCAATTCAGGCTTTTAT 1200
401 ThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProPro 420

```

```

Db      1201  |||||TAGAAGCGGATGGGATATAAAATGCAACCGTTATGTGTTATTAAGACCGACTCTCTCC 1260
QY      421  GlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheLeuLysGlyArgAsn 440
Db      1261  GGAGAAATTTTCATCCCGCGGATACAGGGACTTTTGAACTTTTATATAAGGGCAGAAAC 1320
QY      441  PheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArg 460
Db      1321  TTTACCCCGACCGCGGGTATTGTGTATAGCGGCGACGAACCCATCATGAACACTGCGG 1380
QY      461  AsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetVal 480
Db      1381  AACTGGTACCGTCAAAACCCGATACACAGCAGCGCTTACACTCGACAAATCAAAATATGGTC 1440
QY      481  IleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyr 500
Db      1441  ATTACCAAGCCCGGCAAAACAAATGGGAACAGGAATTAACCTTGATGTGCTTACCTAT 1500
QY      501  ThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLys 520
Db      1501  ACCAACCCAGCTATCGAATCTGGACCATCAGCGCAGTGTACTTTTCATCAACAAAAA 1560
QY      521  TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrp 540
Db      1561  TACTTTCTGTCATCGATAGGCAATAGCGAAGCTACCGGAACCTGGCGGTACACTGG 1620
QY      541  GlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrThr 560
Db      1621  CAGCTTAAAGAGACAGCAACCTGTTTTCGATAGACAAAGAACCGGGTTTACACCACT 1680
QY      561  TyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn 580
Db      1681  TACAGAGATGTTAAACAACTGATGATCCATCGTTGAATCGGACAGGACCGCTCAAT 1740
QY      581  GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal 600
Db      1741  GAAGAAGAAGGAAAGGTACTTATGTTTACAAATAGAGAGCTGAAAGACCTGCTTCGTA 1800
QY      601  PheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyr 620
Db      1801  TTTGAAAGCCCTAAAGAAAGTACCGGCACACAAAAATTTGTAGTATAGTTTATCCATAC 1860
QY      621  AspGlyGluLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLys 640
Db      1861  GACGGCCAGAGGCTCCAGAGTACGATACGGAACCAAGGCAATGATTTTGAGAAA 1920
QY      641  GlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnGlnLeuValLeuValPro 659
Db      1921  GGCAAGCTTAACTTAACCTTACCAATTAAGGAAACACAGCTTGTGTGTTCTCT 1977

RESULT 2
I71365
LOCUS      I71365
DEFINITION Sequence 3 from patent US 5681733.
ACCESSION I71365
VERSION   I71365.1 GI:3007500
KEYWORDS
SOURCE
ORGANISM  Unknown.
REFERENCE  1 (bases 1 to 1980)
AUTHORS   Su,H., Blain,F., Bennett,C., Gu,K., Zimmermann,J. and Musil,R.
TITLE      Nucleic acid sequences and expression systems for heparinase II and
            heparinase III derived from Flavobacterium heparinum
JOURNAL    Patent: US 5681733-A 3 28-OCT-1997;
FEATURES   Location/Qualifiers
            1..1980
            /organism="unknown"
BASE COUNT 629 a 432 c 439 g 480 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 1 85e-280 Length: 1980
Score: 3494.00 Matches: 659
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

```

US-09-802-285A-2 (1-659) x I71365 (1-1980)

```

QY      1  MetThrThrLysIlePheLysArgIleIleValPheAlaValIleAlaLeuSerSerGly 20
Db      1  AUGACTACGAAATTTTAAAGGATCATTTGTTGCTGTAATTCCTATCGTCGGGA 60
QY      21  AsnIleLeuAlaGlnSerSerIleThrArgLysAspPheAspHisIleAsnLeuGlu 40
Db      61  AATATACCTTGCACAAAGCTCTTCATTACCGAGAAAGATTTTGACCACATCAACCTTGAG 120
QY      41  TyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAspAlaAla 60
Db      121  TATTCCGGACTGGAAAAGGTTAATAAGCAGTTGCTGCCGCAACTATGACGATGCGGC 180
QY      61  LysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsn 80
Db      181  AAAGCATTACTGTCATACACTACAGGGAAGAAAGTAAGGCCAGGAACTGATTTCAATAAT 240
QY      81  AlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAsp 100
Db      241  GCAGAAAGCCCTGCCGATATACGCCAGCCCATAGATAAGGTTACGCGTCAAAATGGCCGAC 300
QY      101  LysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLys 120
Db      301  AAGGCTTTGGTCCACCAAGTTCAACCGCAAGGCTACGGCTATTTTATGTTGATAA 360
QY      121  AspIleAsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArg 140
Db      361  GACATCAACTGCGAGATGTGGCGGTAAAGACAATGAAGTAGCTGCGTGCGAGTTGACCGT 420
QY      141  ValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla 160
Db      421  GTAAATGTGTCAGCCTATGCGCTGTTTATCAGCTACGCGGCGATGAAAAATATGCA 480
QY      161  ArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGln 180
Db      481  AGAGAATGGGTATATCAGTACAGCAGATTGGCCAGAAAAACCCATGCGGCTGTGCGCAG 540
QY      181  AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro 200
Db      541  GATAATGATAAATTTGTGTGCGCGCCCTTGAAGTGTGCGACAGGGTACAAGTCTTCCC 600
QY      201  ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220
Db      601  CCAACCTTCAGCTTATTGTAAACTCGCCAGCCTTTTACCCAGCCTTTTAAATGGAATTT 660
QY      221  LeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsn 240
Db      661  TTAACAGTTTACCAACAGCGGCCGATATTATCTACGATATTGCGGAAACAGGGAAC 720
QY      241  HisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys 260
Db      721  CACCGTTTATTGGAAGCCCAACGCAACTGTTTGCAGGGGTATCTTCCCTGAAATTA 780
QY      261  AspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleLysGln 280
Db      781  GATTCCACCAAGATGAGGCAACCGGCATATCGGTGCTGAACACCGCAGATCAAAAACAG 840
QY      281  ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAsp 300
Db      841  GTTTATGCCGATGGGATGCAGTTTGAACCTTTTACCATAATTACATGTAGTCCCATCAT 900
QY      301  IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln 320
Db      901  ATCTTCTTAAGGCCCTATGGTTCTCCAAAACGAGTTTAACCTTGAAAAAGAAATTTCCGCA 960

```

```
QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
DB 961 TCTATGTCACAACTGTGAAAATATGATTAATGCGCTGATCAGTATTTTCACGCGCAT 1020
QY 341 TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGln 360
DB 1021 TATAACACCCCTATGTTGGAGATTTCATGATTACAGATATAAAATTTTCAGGATGCACAG 1080
QY 361 PheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAsp 380
DB 1081 TTTGCCAGCTGGCCCGGTTTCCCGCAACACCGGCATATAAAATATTTTGCTACAGAT 1140
QY 381 GlyLysGlnGlyAlaProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyr 400
DB 1141 GGCACACAGGTAAAGCGCTTAACCTTTTATCCAAGCATTCAGCAATGCAGCTTTTAT 1200
QY 401 ThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProPro 420
DB 1201 ACGTTTAGAAGCGGATGGGATAAAATGCAACCGTTATGGTATTAAAAAGCCAGTCTCC 1260
QY 421 GlyGluPheHisAlaGlnProAsnGlyThrPheGluLeuPheIleLysGlyArgAsn 440
DB 1261 GGGGAATTTTCATGCCCGGATACCGGACTTTTGAACCTTTTATAAGGGCGAGAAC 1320
QY 441 PheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArg 460
DB 1321 TTTACCCAGAGCGCGGGTATTTGTGTATACGCGGAGCCATCATGAACTGGG 1380
QY 461 AsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetVal 480
DB 1381 AACTGGTACCGTCAAAACCGCATACACAGCAGCTTACATCGACCAATCAAAATATGGTC 1440
QY 481 IleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyr 500
DB 1441 ATTCAAAAGCCCGCAAAACAAATGGGAAACAGGAAATAACCTTGATGTGCTTACCAT 1500
QY 501 ThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLys 520
DB 1501 ACCAACCCAAAGTATCCGAATCTGGACCATCAGCGCAGTGACTTTTCATCAACAAAAA 1560
QY 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrp 540
DB 1561 TACTTTCTGGTCATCGATAGGCAATAGCGGAAGCTACCGGAAACCTGGCGGTACATCG 1620
QY 541 GlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrThr 560
DB 1621 CAGCTTAAAGAAAGCAGCAACCTGTTTTCGATAAGACAAAGACCGGTTTACACCAT 1680
QY 561 TyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn 580
DB 1681 TACAGAGATGGTAAACACCTGATGATCCAATCGTTGAATGCGACAGGACGACCTCAAT 1740
QY 581 GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal 600
DB 1741 GAAGAAGAGGAAAGAGTATCTTATGTTTACAATAAGAGAGCTGAAAGACCTGCTTCGTA 1800
QY 601 PheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyr 620
DB 1801 TTTGAAAGACCTTAAAGAAAGATCCCGGCACACAAATTTTGTGATATAGTTTATCCATAC 1860
QY 621 AspGlyGlnLysAlaProGluLysSerIleArgGluAsnLysGlyAsnAspPheGluLys 640
DB 1861 GACGCCCAAGAGCTCCAGAGATCAGCATACGGAAGAAACCAAGGCAATGATTTTGAGAA 1920
QY 641 GlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnGlnLeuValLeuValPro 659
DB 1921 GGCAAGCTTAATCTAACCCCTTACATTAACGGAAACCAACAGCTTGTGTGTTGTTCT 1977
RESULT 3
AE014145/c 50327 bp DNA linear BCT 19-JUL-2002
LOCUS AE014145 Streptococcus pyogenes MGAS315, section 10 of 37 of the complete
DEFINITION genome.
```

```
ACCESSION AE014145 AE014074
VERSION AE014145.1 GI:21904143
SOURCE Streptococcus pyogenes MGAS315.
ORGANISM Streptococcus pyogenes MGAS315
KEYWORDS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 50327)
AUTHORS Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
TITLE Genome sequence of a serotype M3 strain of group A Streptococcus:
Phage-encoded toxins, the high-virulence phenotype, and clone
emergence
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
REFERENCE 2 (bases 1 to 50327)
AUTHORS Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
Hamilton, MT 59840, USA
FEATURES
source
1..50327
/organism="Streptococcus pyogenes MGAS315"
/strain="MGAS315"
/serotype="M3"
/db_xref="taxon:198466"
/note="group: A"
95..307
/gene="SpyM3_0419"
95..307
/gene="SpyM3_0419"
/note="best non-GAS blastp hit: gb|AAK74919.1| (AE007385)
conserved hypothetical protein [Streptococcus pneumoniae
TIG84], and gb|AAK99494.1| (AE008446) Conserved
hypothetical protein [Streptococcus pneumoniae R6]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAW79026.1"
/db_xref="GI:21904144"
/translation="MAEFTNIEEHLTLSENDKGTWTKELNRSNGAEAKWDIRTS
PDHFMGIGITLTNEEFTILDAPKK"
complement(686..2179)
/gene="lyss"
/note="SpyM3_0420"
complement(686..2179)
/gene="lyss"
/note="best blastp hit: gb|AAK33574.1| (AE006515) putative
lysyl-tRNA synthetase [Streptococcus pyogenesM1 GAS]"
/codon_start=1
/transl_table=11
/product="putative lysyl-tRNA synthetase"
/protein_id="AAW79027.1"
/db_xref="GI:21904145"
/translation="MSNQHIEELNDQIVREKMMALAEQQIDPPGKFRGTANSDEL
KENYADKTKEELHNEITAIAGRLMTXRGKVGFAHLQDREGQIQIYVRKDSYGED
NYEIPKKADLGDFTGVEGVMRTDMGELSIKATKLHLSKSLRPLPERFHLGIDLETI
YKRHLDIISNRSFDRFVTRSKMSEIRRYLDGLDPLEVETPLVHNSAGGAAPFV
THNAQNDLMLRLATLHLKELVGGMERYEIGRIFRNMGDMATHNPEITSIVYQ
AYAYLDIMNLTEGLIQHAAKAVKGDGIDYQGTETIRINEPFRKRVHMDA.IKENVGVD
FWPENTVEEALAKQVPLEKFTSVGHILINAFEEFVEELIQVTFVFGHPVEVS
PLAKNPEDTFTDFELFMKTYANAFTEINDPFDQLSRFAQAQAQKELGDDSEATG
IDYDFVEALRYGMPPTGGLGIDRLCMLLTNTTTIRDLVLLFPTMKP"
2353..3255
/gene="SpyM3_0421"
2353..3255
/gene="SpyM3_0421"
/note="best non-GAS blastp hit: emb|CAC98713.1| (AJ591976)
```



```

lmo0635 [Listeria monocytogenes]
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA079028.1"
/db_xref="GI:21904146"
/translations="MITAIVFDVDDTIYDQQAQYRIAMEKCFDPDMSVMNOQVIRPR
HYSDVGPFRVMAGEWTEYFRWCKETLLEFGYREIDRAAGVHFOEYVEHELENITM
LDWRMTLDPLSKNVPKMGILITNGPEHQHKKVRKGLYIYIDPKRVIYSQAIGFQKP
EXEIFNLAAEQFMNFOITLIVGSDYNDIMGAPGWSHMFNHRGRQKPKGTPIY
DVAIDNPEQIFGAVKVLFDLPDKNFIPDINDKKNPILQMGINNGLAMAAERLLESNS
IDKWTLLRLTKQEKVIRLKYAR"
/complement(3363..3986)
/gene="SpvM3_0422"
/complement(3363..3986)
/gene="SpvM3_0422"
/notes="best blast hit: gb|AAK33576.1| (AE006515) putative
phosphoglycerate mutase [Streptococcus pyogenes M1 GAS],
possible pseudogene aligns with N-terminal portion (amino
acids 1-207) of PgmA"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA079029.1"
/db_xref="GI:21904147"
/translations="MKLYFVRHCKTLWNLEGRFQAGGDSPLLEAKDETHLLGKELA
KVAFDAVYSDIORAMATAAILDADFQPKLYHTQRLREWRGLKLGKAGKIATMAIY
PQOMLAFRENLAQFKEDQFAESIYQIQRVCHLIQSFKDXYONVLIIVGHGANLTAT
IRSLGFEPALLAKGLDNTSITILETKDYLYDCLINWDSFIEV"
/complement(4327..4806)
/gene="SpvM3_0423"
/complement(4327..4806)
/gene="SpvM3_0423"
/notes="best non-GAS blast hit: gb|AAK75104.1| (AE007402)
regulatory protein, putative [Streptococcus pneumoniae
TIGR4]"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AA079030.1"
/db_xref="GI:21904148"
/translations="MAKTKLKKTLVEQLDKNIAHQGLKLNALRGDPPDILQPSDI
KYTLALTGQTLGLIPLTLEHSEKQAKVGNKXSVMPKQDLQKTYTHGANN
PVGIRQKHSYPFIDQITALEKGQIIVSAGEVGRSIIKSQALADFVGSFADLKKRK"
/complement(4896..5459)
/gene="SpvM3_0424"
/complement(4896..5459)
/gene="SpvM3_0424"
/notes="best non-GAS blast hit: emb|CAC99507.1| (AL591979)
similar to unknown proteins [Listeria monocytogenes]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA079031.1"
/db_xref="GI:21904149"
/translations="WSPNTNKKYLIEAALFAALMALSFIPDPAGWFSYGAIALVI
FSLRLKGLMAGLWGLHVLGVKGVYLSQVFEIYLAFTSGLGAGSDSLIK
TLRRQOTFAVLAIVASLAVTVRLVHFLAGLIFWGSYAPGSAVWYSPVNGTA
GVLTFLITCLAMIALPIHPQLFDPKD"
/complement(5728..6576)
/gene="SpvM3_0425"
/complement(5728..6576)
/gene="SpvM3_0425"
/notes="best blast hit: gb|AAK33579.1| (AE006515) putative
endolysin, phase associated [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="putative endolysin, phase associated"
/protein_id="AA079032.1"
/db_xref="GI:21904150"
/translations="WRRKIKPIVILVFFILLAMVLIIGKROANHAQKEVEDAKSHIP
IATSNPSKAKTSTETEDFILNPIDVSCWQLPEEIDYDLSRHISGAIVRYGSGQI

```

```

TAHNAFTTIDKSPKTHIKERQKNVPVAVYSYALGRSTKEMKEBARAFYKNAAPY
NPYYIWDVEEATKMDKNKGVTAFREELKGLGAENGLYIGTYFMAEQDILSTKGFDAV
WIPTGSDSGSYEAAPNTTLDYDLHQYTSQYLSGFNNALDLNQIAVTOKTKTFEKL
FGNSNN"
/genes="SpvM3_0426"
/genes="SpvM3_0426"
/notes="best non-GAS blast hit: gb|AAK05270.1|AE006349_5
(AE006349) unknown protein [Lactococcus lactis subsp.
lactis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA079033.1"
/db_xref="GI:21904151"
/translations="MSAKFMSRIAMSALCVLRVMVFISLPNVQVPTAFLLSYLLVFG
LAEAVLVMCLFLSAPLLGFGPWFQVTCFVLILLWFVLYPLSQQPKYQLGCO
AEIVALCGLLYGLVIDTCFAYLYSMFWSYVLGMPFNIAHALSTLVFFVFWMLFRR
LISEQKI"
/genes="SpvM3_0427"
/genes="SpvM3_0427"
/notes="best non-GAS blast hit: gb|AAK05269.1|AE006349_4
(AE006349) unknown protein [Lactococcus lactis subsp.
lactis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA079034.1"
/db_xref="GI:21904152"
/translations="MNRKFTGFLAVLMLVACSQGTQOTQTPSPVKADHHVRLVVK
BNTNIDKXVSGKGTIVLEILKNDYEVKEDGFIADIGIEQDTKANKYWLKVGNG
MADKADQITVKGDSISFVQEVK"
/complement(7817..8296)
/gene="SpvM3_0428"
/complement(7817..8296)
/gene="SpvM3_0428"
/notes="best blast hit: gb|AAK33582.1| (AE006515) putative
glutathione peroxidase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative glutathione peroxidase"
/protein_id="AA079035.1"
/db_xref="GI:21904153"
/translations="MNPVLYDFTVKAQDSDISLVNYKEKVLVWNTATKGLTPQYOA
LQALYDTHDKGFVLDPPCNQFLNQTGCEAEINRFCSLYHTTTPERAKIKVNGKD
ADPLFTWLKEKSGPLGKRIEWNFKFLIDQGVQIKRYSKTDPKLIEDLKALLG"
/complement(8289..10088)
/gene="pepf"

```

Alignment Scores:

Pred. No.:	Length:	Matches:	Score:
2,63e-08	50327	235.50	152
Percent Similarity:		Conservative:	87
Best Local Similarity:		Mismatches:	282
Query Match:		Indels:	140
DB:		Gaps:	26

US-09-802-285A-2 (1-659) x AE014145 (1-50327)

QY 64 LeuAlaTyrrArgGluLysSerLysAlaArgGluProAspPheSer----- 79
 |||||:::|
 28575 CTTGCTTTCGCACGCTTTAAAGAAACAGTC---AATCCAGACTTTTGCAGAAATTACCTG 28519
 |||||:::|

QY 80 -----AsnAlaGluLysProAlaSplleArgGlnProIleAspLysValThrArg 96
 |||||
 28518 CTAGACTATCAGGCTAACAGCTATGCTGAT-----CAAAAA 28483
 |||||

QY 97 GluMetAlaAspLysAlaLeuValHisGln-----PheGlnProHisLysGlyTyrGly 114
 :|||:::|
 28482 AGAATAGCTGACCTACTCTTAACCAATACCTTTTATTGAGATAATTG-GGATATGGA 28424
 :|||:::|

QY	115	TyrPheAspTyrGly-----Lysasp-ileAsnTrpGlnMetTrpProValIysAs	131
Db	28423	GCCTGCTCTATTCTTATCGTTTGGAGACTATCACTTGGCAAAAGCGCGTGACAGATGA	28364
QY	131	pAsnGluValArgTrpGlnLeuHisArgValIysTrpGlnAlaMetAlaLeuValIy	151
Db	28363	TCCTGAGTGAACCTTATCTTAATCGACAGACCTACCTCCAAACATTAACTCTCGTCTA	28304
QY	151	zHisAlaThrGlyAspGluLysTyr-----AlaArgGluTrpValTyrGlnTyrSe	168
Db	28303	CCTGTGTGAAGAGATGAACGGTATCTCTTAACAGCTAAAGGTTTCATTCTCAATTGGAT	28244
QY	168	rAspTrpAla---ArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTr	187
Db	28243	TGAATCAGCTATTCTTTAGACCTTAAGGGTTAGCTACTAG-AACCTTTAGATAC---	28188
QY	187	pArgProLeuGluValSerAspArgValGlnSerLeuPro-ProThrPheSerLeuPheV	207
Db	28187	GCATCCGTGTGTTTGTGGGTCAAATGTTTGATTACCTTAATCTTTTAAAC-----	28135
QY	207	AlaAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnG	227
Db	28134	-----GCCCTTAACCAACCAAGAGGAAGCCTCATCTAGCTTCCCTGGAAAAAC	28086
QY	227	InAlaAspTyrLeuSerThrHisTyrAlaGluGln-----GlyAsnHisArgLeuP	244
Db	28085	AACTTCATTTCTCCATACCACTACTCTGTATATATACAGCCTCAGCAATTGGGNAATCT	28026
QY	244	heGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheIysAspSerProA	264
Db	28025	TGCAAAACCACTCGATTATTTATAGCTGATGCTACTACTGTGTTGACCTTGATATTGCTG	27966
QY	264	rgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleIysIysGlnValTyrAlaA	284
Db	27965	CAGCAACACCTTTCGCTAGAAAAGAAATTCACCAACAAATCGCACTACCAAAATTCCTGAAG	27906
QY	284	spGlyMetGlnPheGluLeuSerProIleTyrHisValAlaLalleAspIlePheLeuL	304
Db	27905	ACGGTAGTCAATTTGACACATCAACCATGTCACCATGTGGAGGTGTAAATCCCTTTGG	27846
QY	304	yAlaTyrGlySerAlaIysArgValAsnLeuGluLysGluPheProGlnSerTyrValG	324
Db	27845	AAATTA-----ACTGCTCTAGTCCCTGATTATCTTCCACAGCTTAGAC	27804
QY	324	InThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnThrP	344
Db	27803	CTACTTTGCTTGCCATGAGTAGTATTATCTCCTCAAAATGATGATCTGACCAACACAGA	27744
QY	344	roMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPheAlaSer-	363
Db	27743	TCCCTTGGAGATAGTAGTATTATCGATACAGAGATATTCTAACACTAGCAGCCACTA	27684
QY	364	-----TrpAlaArgValPheProAlaAsnGlnAlalleLysTyrP	377
Db	27683	TCCTTGAGAACCAACACTTAAAGTGTGCTTTTCCACACTTGAT---ATCGATAGTC	27627
QY	377	heAlaThrAspGlyLysGlnGly-----LysAlap	387
Db	27626	TTCTCCTTCTAGAGAAAAGGTGTCACACTTTTGAACAGTTTACCATTCCAAACCTTAC	27567
QY	387	roAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTrpA	407
Db	27566	CAACTTTT---GGCGATCACTTTGAACATTCGGGGACATTAACATC-----A	27522
QY	407	spLysAsnAlaThrValMetValLeuLysAlaSerProProGlyGluPheHisAlaGlnP	427
Db	27521	ATCAAGAAAATTTATTTATCTTTTTCAAAATATGGGCCATCGTAGTTCCCATACCCATA	27462
QY	427	roAspAsnGlyThrPheGluLeuPheIleIysGlyArgAsnPheThrProAspAlaGlyV	447
Db	27461	GGACCAAAATAGTATCTGCTCTATATATAGGACCAACCACTTTTCTGTGATGCTGGGC	27402
QY	447	alPheValTyrSerGlyAspGluAlalleMetLysLeuArgAsnTrpTyrArgGlnThrA	467
Db	27401	GATACACTTACAAAGAGAA-----CCATACGCTATGCCCTTAAAGCGCCA	27354
QY	467	rgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnA	487
Db	27353	GCCATCAITCTACCGCTTCTTAGAAGAACAA-----TTACCTGAGCAGATTGATA	27303
QY	487	snLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnProSerTyr---	505
Db	27302	GCTCTGGGCT-----TACCTTTCTTACCCCAATTCCAATTATGTC	27261
QY	505	-----	505
Db	27260	ATCTTCGACAAATCGACACGCTTTACTTTATTGATGGTAGCTACACAGCGCAATTTCTGT	27201
QY	506	-----ProAsnLeuAspHisGlnArgSerValLeuPheIleAsnIysIysTyrPheLeuV	524
Db	27200	ATAGAAACAACACTTACCAACATGACCGCAAAATTTCTCTTACCACAGGAATTTTCTTA	27141
QY	524	alIleAspArgAlalleGlyGluAlaThrGlyAsn-----L	536
Db	27140	TCATTTGATACGATA-----CAGGCTCAAGGCAACCACTGTTGTGAGTCGAATTTATTC	27087
QY	536	euGlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnA	556
Db	27086	TTGATAATCACTTAGAAGTAAGACAGAT-----CACCTTAGTGACCTCA	27042
QY	556	rgValTyrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspA	576
Db	27041	GACTAATCAGTGACTAC-----	27025
QY	576	rgThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysA	596
Db	27024	-----CCTTCTACTATAGAAAGACGATTTTGTCAAAAATAACACCACTACCTGACTA	26970
QY	596	rgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerI	616
Db	26969	GTCACAACTCATCAACGAAACCTTTTAAAGACAGAGGTTCACCACT-----ACCC	26916
QY	616	leValTyrProTyrAspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyA	636
Db	26915	TGTTAGTCCCTGACGATACCAAAAGTTACTCCCTTAACACCACTTCAACAGCAACGCA	26856
QY	636	snAspPheGluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGln	653
Db	26855	ACCAATTGAGACCGCC-----CTTCTCGCATTTAAAGAAAGCAG	26812
RESULT 4			
LOCUS	AE006517/c	12195 bp	DNA linear BCT 01-JUN-2001
DEFINITION	Streptococcus pyogenes M1 GAS strain SF370, section 46 of 167 of the complete genome.		
ACCESSION	AE006517	AE004092	
VERSION	AE006517.1	GI:13621815	
KEYWORDS			
SOURCE	Streptococcus pyogenes M1 GAS.		
ORGANISM	Streptococcus pyogenes M1 GAS.		
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
REFERENCE	1 (bases 1 to 12195)		
AUTHORS	Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K., Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.		
TITLE	Complete genome sequence of an M1 strain of Streptococcus pyogenes		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)		
MEDLINE	21132684		
PUBMED	11296296		
REFERENCE	2 (bases 1 to 12195)		
AUTHORS	Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K., Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J.,		

Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
 Direct Submission
 Submitted (10-APR-2001) Department of Microbiology and Immunology,
 University of Oklahoma Health Sciences Center, 940 SL Young Blvd.,
 Oklahoma City, OK 73104, USA
 Location/Qualifiers
 1. .12195
 /organism="Streptococcus pyogenes M1 GAS"
 /strain="SF370"
 /serotype="M1"
 /db_xref="taxon:160490"
 complement (106. .339)
 /gene="SPY0619"
 complement (106. .339)
 /gene="SPY0619"
 /note="Best Blastp hit = no blast hits"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK33591.1"
 /db_xref="GI:13621816"
 /translations="MEFYALFLSLGSKFRVTSPLVLTPLLPOQKAVALAHVTRNNKK
 SVKSAEQMLFEHQTFGVSIILRLKKEYFLQSP"
 complement (411. .1712)
 /gene="SPY0621"
 complement (411. .1712)
 /gene="SPY0621"
 /note="Best Blastp hit = sp|P39651|YWFO_BACSU HYPOTHETICAL
 51.0 KD PROTEIN IN PTA 3'REGION >gi|7450837|pir||G70056
 yWfo protein - Bacillus subtilis
 >gi|1561567|emb|CAB02494.1| (Z80355) Unknown [Bacillus
 subtilis] >gi|2636296|emb|CAB15787.1| (Z99123) alternate
 gene name: ipa-93d
 similar to hypothetical proteins [Bacillus subtilis]"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK33592.1"
 /db_xref="GI:13621817"
 /translations="MNEKFRDVRVNYIHIDNPLIYDLINTKEFQRLRIKQVPTTAF
 THGASHRSFLCGVEIATRTVTAIFEEKYADIMNKDESVMTMAALHDIHGAYS
 HPEVFLDHFAETGEIITNPETEINAILVRHAPDFPKQVASVINHTYPNKOVVOLI
 SSQICDRMDYLLRDSYFAANYGQFDLRILVRIPVBDGIVFHSGHMVAVEDYIVS
 RFQMVQVYFHPASRAVELLQNLAKRAOLYEQQAYFQKTAAGLIPFEKKANLAD
 YIALDDGVNTYFQVMASEDHILSLASEFNRLKLSVTFDQDSQGLERLRQLVE
 SVGDPDYTGTHINDELDPYDVRPLENPRQIENMQDGSIAELSQLSPIVKAULTG
 TYGDRRRFFPKMELEDDLFPASKETFMYSYISNGHFHSQ"
 1794. .2180
 /gene="SPY0622"
 1794. .2180
 /gene="SPY0622"
 /note="Best Blastp hit = no blast hits"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK33593.1"
 /db_xref="GI:13621818"
 /translations="MKLQTNHIRFGDETEIIQEIHDCEWREKGGVQYLIYQNTDKEK
 VIKNETELTWSRFSNPQIMKFFAGKRVLIATLPTMGVQQPFLDITSHYLDSCQK
 LDLYHLLQAQTEMLFASYHLELSD"
 2411. .5092
 /gene="pacL"
 /note="SPY0623"
 2411. .5092
 /gene="pacL"
 /EC_number="3.6.1.-"
 /note="Best Blastp hit = pir||H69877 calcium-transporting
 ATPase homolog yloB - Bacillus subtilis
 >gi|2337795|emb|CAA74269.1| (Y13937) putative PacL protein
 [Bacillus subtilis] >gi|2633938|emb|CAB13439.1| (Z99112)
 similar to calcium-transporting ATPase [Bacillus
 subtilis]"
 /codon_start=1

/transl_table=11
 /product="putative calcium-transporting ATPase"
 /protein_id="AAK33594.1"
 /db_xref="GI:13621819"
 /translation="MSKEQRHEAFYTSQETVLAQLETSGELTSAQAKERLAAYGRN
 ELDEGRKSLFMKLFQFDLMIILIIAALISLVLEGMEGLTDAIILAVILNRAAP
 GYQEQAAEAIALKSMSPLEARIRRDGHVTEIDSKELVPGDIVILLEAGDVPADUR
 LEANSIKIEARALTGESVFVEKDSVAGISGDRVNMGYONNVTVGRGIVGIT
 NTGMYTEVGIAGMLANADETPTLFKONLDNLSKILTYAILVIAAVTFAYVFLRGQH
 PLEGLTSMVALVAAIPEGLPAIVTSLIGTQVLAKRNAIRKLPKAVETLGSIEIILA
 SDKTGLTMNVTKEVYNTNGTSSADIAFDNTTLRVNMFANDTKVDPSGKLIIGDP
 QETALVEGLDRNFVREMAVEPRVAELPEFSDRKLMSITHKQAGQFIAPVKGADP
 QLLKRVQLENGQIRPITDADKTIILDTNKSILAKALRULMAYKYSALPTEIETI
 VEANLVFSLGVGMIDPERPEAAQVAKAEIRPIMITGDHQTDAKAKRGLIIEE
 GDVDHVTGAELNELSDETFQVFOYSVYARVSPHKVRIKVAWEGKVVAMTGGG
 VNDPSLTADIGIGMITGTEVSGKASMDVADNNFATIIVAWEGKVFVSIQKTI
 QYLLSANMAEVTIFLATLFGWDLQVHLLWNLVMTDTLPALGVPAPGVMKHK
 PRGRKSSFFDGGVKRAILYQGAFTILVGVGVFALMFPEHTSYHDVHADALTMAYVT
 LGLIQLAHAYNVKSVYQSIFTVGLPKNKLPNYSIPVAFVAMATVTVFPGNQFFHVT
 LITOMLVIIIGSLMVLVVELVKAQVRSLSGQDEKAI"
 complement (5176. .6171)
 /gene="regR"
 /notes="SPY0627"
 complement (5176. .6171)
 /gene="regR"
 /note="Best Blastp hit = sp|P50844|KDGR_BACSU KDG OPERON
 REPRESSOR >gi|7442871|pir||B69648 transcription repressor
 of pectin utilization operon kdgr - Bacillus subtilis
 >gi|1146188|gb|AAB38478.1| (U47838) repressor protein
 [Bacillus subtilis] >gi|2634631|emb|CAB14129.1| (Z99115)
 transcriptional regulator (LacI family) [Bacillus
 subtilis]"
 /codon_start=1
 /transl_table=11
 /product="putative transcriptional regulator (LacI
 family)"
 /protein_id="AAK33595.1"
 /db_xref="GI:13621820"
 /translations="MQRKVTIKIDIAELAKTSKTTVSFYLNGRFDMSEETKNRISSEI
 KATNYSFARSNAKSLIGVIGDITNSFNQIVKGIESKAEQFYQIIIGNSN
 YDPSDEDELTENKLDVGVGFIQPTNSPKYSRIIDIKKKVYVDFDOLYHRTWV
 KTNVDVDEITQQCICDKGYEHFIMTGNPNLLSTFIEPASGFIDVLEAHNLTHOMI
 IDENOTSRAIAQFQGLSITKSLVFPNCWALPKVFTAMKSLIKFNIPEIGLVGFNI
 EWKFSPLTITIIIPAVEGEQATKILIDIDIEGHSQEAQKQIFDCQVNWQESTP"
 complement (6235. .8142)
 /gene="SPY0628"
 complement (6235. .8142)
 /gene="SPY0628"
 /note="Best Blastp hit = pir||T36084 hypothetical protein
 SCE134.02c - Streptomyces coelicolor
 >gi|4678627|emb|CAB41200.1| (AL049661) hypothetical
 protein [Streptomyces coelicolor A3(2)]"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK33596.1"
 /db_xref="GI:13621821"
 /translations="MFFAFAREKTVNPDFCRNYLLDYQTSYADQKRIADLLLTNTF
 LPEDNMEPCHIPYHLDPIITWQEAVIDDPEWNFMNLNRTYLOKLIILVYLDERYL
 LTAQGFILNWIESAIPDQGLATRLDTGIRCFAWVKCLIVLNLNALTQKESLIL
 ASLEKLOPLHANVLDKYSLNWGILOTTAILADAYFGSDLDIAAATAPAKELTQQ
 TALQILEGDSQFEQSTMVHVEVLKALLELTALVPDYLPOLRITLAMSYLDKMTGPD
 HKQIFUGSDVDTRDITLQANLILEBHLKAAAFPTDIDISLLLGEKGVHFTQLP
 VQTLPTFAHFEHSGHITINQENYILFFKNGIPGSHSDQNSLCLYKGGQFLCDA
 CRYTYKEEPLRYALKASHHSTAFLEQPEQIDSWAYLSYKPNCHLNRGHVYF
 TEGSYQTOFSDRNNYCHDRQILILPGGFIILDTIQACQNHCLVSOQFILDNDVKTQ
 HSLDRILSDCPTTEETILSKYNOYLTSKHLIKRPFKDKGCTSTLLVPDQKVP
 LTPLOTGRNPETALSMLKGGKQFDYSICVLQEDLIKGEKLVILNLSHKKIRGVVIN
 HITNIIIRLKH"
 complement (8229. .9050)
 /gene="agaD"
 /note="SPY0629"
 complement (8229. .9050)

```

/ gene="agad"
/ note="Best Blastp hit = sp|P42911|PTPD ECOLI PTS SYSTEM,
N-ACETYLGLACTOSAMINE-SPECIFIC IID COMPONENT (EIID-AGA)
(PHOSPHOTRANSFERASE ENZYME II, D COMPONENT)
>gi|7434458|pti|H65103|pts system,
n-acetylglactosamine-specific IID component - Escherichia
coli (strain K-12) >gi|1789529|gb|AAC76174.1| (AE000395)
PTS system, N-acetylglucosamine enzyme IID component 1
[Escherichia coli] >gi|8895756|gb|AAF81092.1|AF228498_12
[Escherichia coli] >gi|8895756|gb|AAF81092.1|AF228498_12
/codon_start=1
/transl_table=11
/product="putative PTS dependent galactosamine IID
component"
/protein_id="AAK33597.1"
/db_xref="GI:13621822"
/translation="MTGSKNLTQKDYLTALRAFFLQNGFNNTYQGIYANVIYPAL
KKHFGNDKRGYLAEDNCFEYNTNPHLPFTSLHLVLENNRPEETRNKLMALMG
PLAGIGDSLQFCLAPLESTIAASLASGLVLGPILFLAMNIILTAIKIGSLGYK
VTSFDKLSQMAVVSRIANVGVTVLAGLAATSVKLTVPITPAAGKVDAAATAQKF
VTIQMLDKIAPALPALPFTLLMYLILKNKWTYKLVILTVIIGVIGSWLGILA"
complement(9037..9819)
/ gene="agaw"
/ note="SPY0630"
/complement(9037..9819)
/ gene="agaw"
/ note="Best Blastp hit = gb|AAF81084.1|AF228498_4
(AF228498) Agaw [Escherichia coli]"
/codon_start=1
/transl_table=11

```

Alignment Scores:

```

Pred. No.: 4, 83e-09 Length: 12195
Score: 234.50 Matches: 140
Percent Similarity: 36.35% Conservative: 85
Best Local Similarity: 22.62% Mismatches: 273
Query Match: 6.71% Indels: 122
DB: 1 Gaps: 23

```

US-09-802-285A-2 (1-659) x AE006517 (1-12195)

```

QY 96 A-rggluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGly--- 114
DB 8046 AAAAGAAAGTGGACCTACTCTTAACCAATACCTTTTATTGAAGATAATGGGATAG 7987
QY 115 -----TyrPheAspTyrGlyLysAsp-----IleAsnTrpGlnMetTrpProValLys 130
DB 7986 GAGCCCTGCCACATCCCTTACCCTTAGATCCCATTTACTTGGCAAGAGGCTGTCTATGAT 7927
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuVal 150
DB 7926 GATCTGAGTGGAACTTATGTTAATCGACAGACCTATCTCAAAAATAATATCTCTGHC 7867
QY 151 TyrHisAlaThrGlyAspGluLysTyr-----AlaArgGluTrpValTyrGlnTyr 167
DB 7866 TACCTTGTGAAAGAGATGAACGGTATCTCTTAACAGCTAAAGGTTTCATCTCAATTGG 7807
QY 168 SerAspTrpAla---ArgLysAsnProLeuGlyLeuSerGlnAsnAspLysPheVal 186
DB 7806 ATGAATACGCTATCTCTTAGACCTTAAGGGTTAGCTACTAG-AACCTTAGATAC--- 7751
QY 187 TrpArgProLeuGluValSerArgValGlnSerLeuPro-ProThrPheSerLeuPh 206
DB 7750 TGGATCCGTTGTTTGGTGGTCAATGTTTGAATTACCTTAATCTTTTAAC----- 7696
QY 206 eValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisG 226
DB 7695 -----GCCTTAACCAAGCAAGAGAAAGCCTCATCTAGCTTCCCTGGAAAA 7649
QY 226 nGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln-----GlyAsnHisArgLe 243
DB 7648 ACAACTTCAATTTCTCCATGCCAACTACTCTTGTATAAATACAGCCTCAGCAATTTGGGAAT 7589

```

```

QY 243 upheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAspSerPr 263
DB 7588 CTTGCAAAACCCTGGATTTTATTAGCTGTAGTCCCTACTTTGGTTCTGACCTTGATATGC 7529
QY 263 oArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluLysLysGlnValTyrAl 283
DB 7528 TGCAGCAACAGCTTTTCGTAGAAAAAATTGACGCAACAAATCGCACTACAAATTCGGA 7469
QY 283 aAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLe 303
DB 7468 AGACGGCGATCAATTGAAACATCAACCATGTATCATGTGTGAAGTACTCAAAAGCCCTTCT 7409
QY 303 uLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrVa 323
DB 7408 GGAATTA-----ACTGCTCTAGTCCCTGATTAATCTCCACAGCTTAG 7367
QY 323 lGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnTh 343
DB 7366 ACCTACTTTGCTTGCATGAGTATTATCTCCTCAAAATGACTGGTCCAGATCATAAACA 7307
QY 343 rProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPheAlaSe 363
DB 7306 AATCCCTCTTGTGTAGATGCGATGTACAGATACCAAGAGATATCTAAACGCTTGCAGCCAC 7247
QY 363 r-----TrpAlaArgValPheProAlaAsnGlnAlaIleLysTyr 376
DB 7246 TATCCTCTGAAGAACCACTTAAAGCTGCTCTTTTCCACACTTGAT--ATCGATAG 7190
QY 376 rPheAlaThrAspGlyLysGlnGly-----LysAl 386
DB 7189 TCTTCTCTCTTAGGAGAAAAAGGTGTCCACACTTTTGAACAGTTTACCCTCCAAACCTT 7130
QY 386 aProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTr 406
DB 7129 ACCAACTTTT---GCGCATCACTTTTGAACATTTCCGGGACATTTACCATC----- 7084
QY 406 pAspLysAsnAlaThrValMetValLeuLysAlaSerProProGlyGluPheHisAlaGl 426
DB 7083 -AATCAAGAAATATTATCTTTTTCNAANAATGGCCCATCGGTAGTTCCCATACCCA 7025
QY 426 nProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGl 446
DB 7024 TAGCGCAACCAATAGTCTCTGTCTATTATTAAGGACCAACCACTTTTCTGTGATCTGG 6965
QY 446 yValPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnTh 466
DB 6964 GCGATACACTTCAAGAAGAA-----CCATTACGCTATGCCCTTAAAGAGCGC 6917
QY 466 rArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGl 486
DB 6916 TAGCCATCATCTACCGCTTCTTAGAAGAACAA-----TTACCTGAGCAGATTGA 6866
QY 486 nAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnProSerTyr--- 505
DB 6865 TAGCTCTCTGGCT-----TACCTTTCTTACCCCAATCCAATTATTG 6824
QY 505 ----- 505
DB 6823 TCACCTTCGACAAAATGGACAGCTTTTACTTTATTGAGGTAGTACCAGACGAATCTTC 6764
QY 506 -----ProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLysTyrPheLe 523
DB 6763 TGATAGGAACAACCTACCAACATGACCGCAAAATCTCATCTTACCACAGGAATTTTCT 6704
QY 523 uValIleAspArgAlaIleGlyGluAlaThrGlyAsn----- 535
DB 6703 TATCATTTGATAGATA-----CAGGCTCAAGGCAACCACTGCTTGTGTGAGTCAATTTAT 6650
QY 536 -LeuGlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAspLysThrLysAs 555
DB 6649 TTTTGTATATCACTTAGACGTAAAGACAGAT----- 6619

```

```

QY 555 nArgValTyrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAs 575
Db 6618 -----CACCITAGTGACCTAGACTATCATGTGA 6590
QY 575 pArgThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsnLysGluLeuTy 595
Db 6589 C---TGCCTTTCACATCAGAGAAACGATTTGTCAAAAAATAACAACCAATACCTGAC 6533
QY 595 sArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSe 615
Db 6532 TAGTCACAACTCAACAAGAAACCTTTAAAGACAAGGGTTCACCACT-----AC 6479
QY 615 rIleValTyrProTyrAspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysG1 635
Db 6478 CCTATTAGTCCCTGACGATACCAAGTACTCCTTAACACCACTTCAACACGCAACG 6419
QY 635 yAsnAspPheGluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGln 653
Db 6418 CAACCCAATTGAGACTGCC-----CTTCTCGCATTTAAAGAAAGCAAGCAG 6373

RESULT 5
LOCUS AE010002/C
DEFINITION Streptococcus pyogenes strain MGAS8232, section 50 of 173 of the
complete genome.
ACCESSION AE010002 AE009949
VERSION
KEYWORDS AE010002.1 GI:19747861
SOURCE Streptococcus pyogenes MGAS8232.
ORGANISM Streptococcus pyogenes MGAS8232
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1 (bases 1 to 12187)
AUTHORS Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.
Genome sequence and comparative microarray analysis of serotype M18
group A Streptococcus strains associated with acute rheumatic fever
outbreaks
Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)
MEDLINE 21927593
PUBMED 11917108
REFERENCE 2 (bases 1 to 12187)
AUTHORS Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.
Direct Submission
Submitted (31-JAN-2002) Laboratory of Human Bacterial
Pathogenesis/Rocky Mountain Laboratories/NIH, 903 S. 4th St.,
Hamilton, MT 59840, USA
FEATURES
source
1. 12187
/organism="Streptococcus pyogenes MGAS8232"
/strain="MGAS8232"
/db_xref="taxon:186103"
complement(106..339)
/gene="spyM18_0686"
complement(106..339)
/gene="spyM18_0686"
/note="best blastp match gb|AAK33591.1| (AE006517)
hypoetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypoetical protein"
/protein_id="AAL97359.1"
/db_xref="GI:19747862"
/translation="MEFVALFLSLGSKSRVTSLPVLTSLPQQAAVLAVHTRRKK
SVVKSAEQPMLFHTGTGFSIURLKKEFLQSF"
complement(411..1451)
/gene="spyM18_0687"
complement(411..1451)

```

```

/gene="spyM18_0687"
/note="best blastp match gb|AAK33592.1| (AE006517)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL97360.1"
/db_xref="GI:19747863"
/translation="MTAALLHDIGHGAYSHTEALFHTDHEAFTQBIITNPETEINAI
LYRVAPDFDKVASVINHTYPNKVQLISSQIDCDMDYLLRDSYFSAAYGQFDLM
RILRVRPVEDGIVPEHSGMAVEDVIVSRFQMYMVYFHPASRAVELILQNLKRAQ
HLYPEQAAVEQKTAPEGLIPPEKKANLADYIALDDGVWNTYFQVWASDELILSLAS
RINRKLKSVTFQDSRGELELRQLVESVGDPDIYTGIIHINFOLDIYRPELEN
PTQLIEMMQDGLSLAELSQLSPIVKALTGTTGYDRFRFPFKEMLELDDLFAFSKETFM
SYISNGHPHFQS
1794..2180
/gene="spyM18_0688"
1794..2180
/gene="spyM18_0688"
/note="best blastp match gb|AAK33593.1| (AE006517)
hypoetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypoetical protein"
/protein_id="AAL97361.1"
/db_xref="GI:19747864"
/translation="MKQLQTNHIRFDETEIIQIHCDEWREKGGYQYLIYQNTDKK
VVIKYNELTMSRFSNPQSIIMKFPAGKKVLIALTPTMGVQQLTDTSHYLDSCQK
LDLHVHLLQAQTEMLFASVHLELSD"
2410..5091
/gene="pacL"
/note="pacL"
2410..5091
/gene="pacL"
/note="best blastp match gb|AAK33594.1| (AE006517)
putative calcium-transporting ATPase [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative calcium-transporting ATPase"
/protein_id="AAL97362.1"
/db_xref="GI:19747865"
/translation="MSKQRHEAFYTQSEETVLAQLETSREGLTSAQAKERLAERYGN
EIDEGEKSLFMKFLDQFKDMLIILVAALSVLTEGMEGLDAIILAVILNAAF
GYQGOAEAAI EALKSMSPILARRDGHVTEIDSKELVPGDIIVLEAGDVPVADLR
LLEANSKLT EEAALTGESVPVEKDLSTAVSEDAGIDRVNMGYSNVTYGRGIVIT
NTGMYTEGVIAGMLANADETTPKLNLDNLSKILTYAILVIAATVAVGVFLRGOH
PLEGLMTSVALAVALPSELPAITVTVVLISLTQVLAKRNAIRKLPVAVETLSTELLA
SDKTGLTNQMTVEKVTNGTLQSSADIAFDNTTLRVNPFANDTKVDFSGKSGDP
TETALVQGLDHNFDVRAMVAEPRAELPFDSDRKIMSTIHKQADKYFIIVKVGAPD
QLLRVTQIEENGQIRPITDAKKTILTNTKSLAKQALRVLMVAYKSDALPTLETEI
VEANLVFGLVGMIDPERPEAAQVAKAEGIRPIMITGDHDTAKALAKRLITEE
DGVDFVTGAEINLSDSEFOKQYVYARVSPEHKVIRIVKQNEGKVVAMTGDG
VNDAPSLATADIGICMGLTGTSEVSKGASMDVADDNFATIIIVAREGRKVFNIQKTI
QYLLSANAEVFTIFALIFGMDVLQPVHLLINLVITDTPALPAIGVSEAPGPKWKHK
PGRKRSFFDGGVKEALYIQGFQFVLVGVGFALMPFETSHYHDHADALTMAVVT
LGLIQLVHAYNVKYSQISFIVGLFPKNLFNYSIEPAFVAFMATVTVVPGFNQFFHVT
LTITQWLVIIGSLMLVVELVKAVQKSLGQDEKAI"
complement(5175..6170)
/gene="spyM18_0692"
complement(5175..6170)
/gene="spyM18_0692"
/note="best blastp match gb|AAK33595.1| (AE006517)
putative transcriptional regulator (LacI family)
[Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL97363.1"
/db_xref="GI:19747866"
/translation="WQRVTKIDIAELAKTSKTTVSFYLNGRFDMSEETKNRIKSTI
EATNVKPSIAARSINAKSKLIGVIGDITNSFNSQIVKGIKSAQEGFYQIIGNSN

```



```
RESULT 6
SPNEU1908
LOCUS
DEFINITION
Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
in ordered pieces.
ACCESSION
AL449930
VERSION
AL449930.1 GI:11545155
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
Streptococcus pneumoniae.
ORGANISM
Streptococcus pneumoniae.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 175936)
Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M.
and Garcia-Bustos, J.F.
Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19F clinical isolate
Microb. Drug Resist. 7 (2), 99-125 (2001)
21335329
MEDLINE
11442348
REFERENCE
2 (bases 1 to 175936)
Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M. and
Garcia-Bustos, J.F.
Direct Submission
Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. .175936
/organism="Streptococcus pneumoniae"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"
BASE COUNT 52994 a 31655 c 38793 g 52447 t 47 others
ORIGIN
Alignment Scores:
Pred. No.: 0.000179 Length: 175936
Score: 198.00 Matches: 140
Percent Similarity: 35.54% Conservative: 96
Best Local Similarity: 21.08% Mismatches: 273
Query Match: 5.67% Indels: 156
DB: 2 Gaps: 25
US-09-802-285A-2 (1-659) x SPNEU1908 (1-175936)
QY 71 SerLysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnPro 90
Db 143534 TCAGAACTATGAGAAAGTCAGAGTTCCTTGACGCTTGATGACAAATCGTTTATG 143593
QY 91 IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis 110
Db 143594 TTTGATAGT--CCTTGGGATAGGAGCCTTGTTCAAAATCCATCAATCCAGCG-- 143647
QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
Db 143648 -----ATGGTATGGGATCAAGTATTTGAAGAT 143674
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuVal 150
Db 143675 GATCCAGATGGCTTATATGCTCAATCGAAGAATATCTCTTGCGATTTATGATAGG 143734
QY 151 TyrHisAlaThrGlyAspGluLysTyrAlaArg-----GluTrp 163
Db 143735 TATGTGGTAGAGGAGATAAGGACTATATTCAAAGTGCAGTCTTCTTCTATTGATGG 143794
```

```
SPNEU1908
LOCUS
DEFINITION
Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
in ordered pieces.
ACCESSION
AL449930
VERSION
AL449930.1 GI:11545155
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
Streptococcus pneumoniae.
ORGANISM
Streptococcus pneumoniae.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 175936)
Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M.
and Garcia-Bustos, J.F.
Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19F clinical isolate
Microb. Drug Resist. 7 (2), 99-125 (2001)
21335329
MEDLINE
11442348
REFERENCE
2 (bases 1 to 175936)
Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M. and
Garcia-Bustos, J.F.
Direct Submission
Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. .175936
/organism="Streptococcus pneumoniae"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"
BASE COUNT 52994 a 31655 c 38793 g 52447 t 47 others
ORIGIN
```

RESULT 6

SPNEU1908

LOCUS

DEFINITION

Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,

in ordered pieces.

ACCESSION

AL449930

VERSION

AL449930.1 GI:11545155

KEYWORDS

HTG; HTGS_PHASE2.

SOURCE

Streptococcus pneumoniae.

ORGANISM

Streptococcus pneumoniae.

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

Streptococcus.

REFERENCE

1 (bases 1 to 175936)

Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,

Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de

Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M.

and Garcia-Bustos, J.F.

Annotated draft genomic sequence from a Streptococcus pneumoniae

type 19F clinical isolate

Microb. Drug Resist. 7 (2), 99-125 (2001)

21335329

MEDLINE

11442348

REFERENCE

2 (bases 1 to 175936)

Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,

Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de

Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M. and

Garcia-Bustos, J.F.

Direct Submission

Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,

Severo Ochoa 2, 28760 Tres Cantos, SPAIN

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

source

1. .175936

/organism="Streptococcus pneumoniae"

/serotype="19F"

/db_xref="taxon:1313"

/clone="G54"

BASE COUNT 52994 a 31655 c 38793 g 52447 t 47 others

ORIGIN


```
QY 164 ValTyrGlnTyrSerAspTyrAlaArgLysAsnProLeuGlyLeuSerGlnAspAsn 183
Db 143795 ATGAGCAGGTGAGAGAAATTTCTCTCAATCTTGATGACTAGAACCTTGATACGGGT 143854
QY 184 -----LysPheValTyrArgProLeuGluValSerAspArgValGlnSerLeuProPro 201
Db 143855 ATTGGTTCCTTACCTTGGTTGAACCTA-----143881
QY 202 ThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeu 221
Db 143882 ---CTCTTGCTCTCTTGAATTTGACTTTGTAGAGGAGAGAACTAGAGAAAATTTTG 143938
QY 222 AsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln-----238
Db 143939 GTCAGTCTAGAAAAGCAGATGACCTTTATGAAAGCCATATCGGCCCAAGTACACCTT 143998
QY 239 GlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGlu 258
Db 143999 AGTAACTGGGGGATTTTACAAACAATTCGATGCTTGCTATCTATTATTCTTTTCAGAT 144058
QY 259 PheLysAspSerProArgTyrArgGlnThrGlyIleSerValLeuAsnThrGluIleLys 278
Db 144059 AAGATGGACCTAGAGAAGCTTACCATTGCTTTCAGAGAGTGTGAACAGCAAAATTTGAG 144118
QY 279 LysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisVal-----296
Db 144119 ACACAGATTTTAGGAGATGAGTCACTTTTGAAACAGTCACTCTCATCATGTAGAGTT 144178
QY 297 -----AlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu 314
Db 144179 TATAAGCCTTACTGGATTGTGTC-----TTG 144208
QY 315 GluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle 334
Db 144209 CTTCCAGATGTCAAGATAGTACCAGAGTGTCTGGAAAGATGGCGACCTATATTCAA 144268
QY 335 SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTyrIleThrAspLys 354
Db 144269 ATGATCAGAGGCTTAGATGACGAACCTTTGGCTTTTGGTGATAGCATCTACAGAAACG 144328
QY 355 AsnPheArgMetAlaGlnPheAla-----362
Db 144329 ACAGAAATTTTGACGCTGTCTGCTGTGTTTGAACAAGAGACCTTCTTAACGCTCTG 144388
QY 363 -----SerTyrAlaArgValPheProAlaAsnGlnAlaIleLysTyrPhe 377
Db 144389 GATGTTAAAGTGAATTTGCTTGTAGCTCTTGTCTGGGGCGAGAAAGGTCAAGCACTG 144448
QY 378 AlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAla 397
Db 144449 CAGGAATTTGAAAGAGAGCTTGGCAGCCTAAG-----TCCATGATCTTTGAAGACTCT 144502
QY 398 GlyPheTyrThrPheArgSerGlyTyrAspLysAsnAlaThrValMetValLeuLysAla 417
Db 144503 GGACATGCTGCTCAATTAAGAT-----GAACATCGTTTATCTATTTTCAAAAAT 144550
QY 418 SerProProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLys 437
Db 144551 GGTCCCTAGAGAGTCCCATAGCCATAGTACGAGAAATAGTTTTTTCCTACGTATCAA 144610
QY 438 GlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMet 457
Db 144611 GGCCAACTATTTTCATAGATGCTGGCGTATTCTTATTCG-GGAGATATATGAACGTTA 144669
QY 458 LysLeuArgAsnTyrTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGln 477
Db 144670 TCTCTTAAAGAGTGTGGAGTCATTCGACCTGCATTGT-----144708
QY 478 AsnMetValIleThrLysAlaArgGlnAsnLysTyrTyrGlu---ThrGlyAsnAsnLeuAsp 496
Db 144709 -----AGATGGAAAGCTCCCGAAGAAATCAGGG 144738
QY 497 ValLeu-ThrTyr-----500
```

```
Db 144739 ATCCTGGGAATATGATACTATCCTCCTCCTGCTTTTGTCCACATAAAGAAAGGAGG 144798
QY 501 -----ThrAsnProSerTyrProAsnLeuAspHi 510
Db 144799 AGTGCAATTATATTGAGGGGCTTATTGGTCAGCAAGAACTGATTTGGCTTATCTT---CA 144855
QY 510 sGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAsp-----ArgAl 528
Db 144856 CAAGAGAAAATCCCTCATGTTGGTAGAGATCTCGGCTCTTGGTAGATGACATCAGGGT 144915
QY 528 aileGlyGluAlaThrGlyAsnLeuGlyValHisTyrGlnLeuLysGluAspSerAsnPr 548
Db 144916 TCAAGTCTAG-----CATGAGGGCTTGACTCAGTTT-----144946
QY 548 oValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGly-----AsnAs 566
Db 144947 -ATCCTTGACAAGGAT-----GTGACCTATCAAGATGGGAAAATCAATCA 144990
QY 566 nLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsnGluGluGluLysVa 586
Db 144991 GTTGAGACTATCGAGT-----GAAGTTGATTTGATTTTGGAGATACCATCAT 145038
QY 586 lSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal-----Ph 601
Db 145039 TTCTCTTAATACAT---GAGCTTGAAAGAGTAGCAAACTCACCAGCGCAATTTCT 145095
QY 601 eGluLysProLysLysAsn-----AlaGlyThrGlnAsnPheValSerIle-----616
Db 145096 TGAGAAATCAGATGCTGGGATATATACCATCATTCGCGATGAGAGATTTTGAATCATCGTCA 145155
QY 617 ----ValTyrProTyrAspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGl 635
Db 145156 TTCTGTCTACAGACAGATCATCGTAGAGTGAATAATGCTTTGGCTTTTGAATGAAANA 145215
QY 635 yAsnAspPheGluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnGlnLe 655
Db 145216 TCACGAAACAGACAAAGCTGATTCTGTATTATTAAGCGAGGATATTCGTGTAGGTGAAAAAT 145275
QY 655 uValLeuVal 658
Db 145276 GTGCCTCGTT 145285

RESULT 7
AE008410 11333 bp DNA linear BCT 13-SEP-2001
LOCUS Streptococcus pneumoniae R6 section 26 of 184 of the complete genome.
ACCESSION AE008410 AE007317
VERSION AE008410.1 GI:15457839
KEYWORDS
SOURCE Streptococcus pneumoniae R6.
ORGANISM Streptococcus pneumoniae R6
          Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
          Streptococcus.
REFERENCE 1 (bases 1 to 11333)
AUTHORS Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,
          DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C.,
          Gilmour, R., Glass, J.S., Khoja, H., Kraft, A., Lagace, R.,
          LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,
          McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicase, T.I.,
          Norris, P.H., O'Garra, M., Peery, R., Robertson, G.T., Rokey, P.,
          Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,
          Zook, C., Baltz, R.H., Jaskunas, S.R., Rostek, P.R. Jr., Skatrud, P.L.
          and Glass, J.I.
TITLE Genome of the bacterium Streptococcus pneumoniae strain R6
JOURNAL J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 11333)
AUTHORS Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,
          DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C.,
          Gilmour, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R.,
```


LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,
McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicast, T.I.,
Norris, F.H., O'Gara, M., Peely, R., Robertson, G.F., Rocky, P.,
Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,
Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L.
and Glass, J.I.
Direct Submission
Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and
Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
FEATURES
source
1. .11333
/organism="Streptococcus pneumoniae R6"
/strain="R6"
/db_xref="taxon:171101"
complement(448. .554)
/rpt_family="RUP element"
/rpt_type="dispersed"
/rpt_unit="Rup A"
complement(758. .1387)
/genes="kdgA"
/note="spr0287"
complement(758. .1387)
/genes="kdgA"
/EC number="4.1.3.16"
/codon_start=1
/transl_table=11
/label="spr0287"
/product="2-keto-3-deoxy-6-phosphogluconate aldolase"
/protein_id="AAK99091.1"
/db_xref="GI:15457840"
/translation="MTKSPDIIELKQKIVAVIRGNTKEGLQASACIKGGIKATEI
AYNQVAGGIIKELVDYDDQSVICGAGVLTAVTARDAILAGANVVPSPHAE
KCNLYSTYIPGCTILTELTWALEAGSEIKLIPGSAISPAVISAKAPIQVSVW
TGGVDLNNIPQWFAAGADAVGICGELKGLASQGNFDRISIAQQYVTLR"
complement(1397. .2398)
/genes="kdgK"
/note="spr0288"
complement(1397. .2398)
/genes="kdgK"
/EC number="2.7.1.45"
/codon_start=1
/transl_table=11
/label="spr0288"
/product="2-keto-3-deoxygluconate kinase"
/protein_id="AAK99092.1"
/db_xref="GI:15457841"
/translation="MTKILLFGEPLIRISPLDATSIGDHVASSTYFGSSEINIA
ALGISTKVTALPANIGDRPLFLFKHQHIDTSSICRLGDRIGLYLENGFGQSEV
FYDRKTSISQIRPNMLDMSLFGQISHFSGITVAIGQVRAILLLLKAKRGI
VVSMDLNLTKMISVLEAKYEFKSFARFTDYCGIDPLMDONLEMPRDSASREV
ENMRLLKAYGFKAIPIHLRSDQDNVQAYALEERFEESVOLKTAVYQRIQSGD
ARTSGALYQLLHSSSKTITIDFAVASATLKCILPGDHLSTATSISNLLANAQDIIR"
complement(2429. .3070)
/genes="spr0289"
complement(2429. .3070)
/genes="spr0289"
/codon_start=1
/transl_table=11
/label="spr0289"
/product="Hypothetical protein"
/protein_id="AAK99093.1"
/db_xref="GI:15457842"
/translation="MKIALINENSQAKNHLYDSLKEATDKDYQFNVGRGE
SQTYYQNGIMAILNTKAVDPVVTGCTGVGAMALNSFFGVCGLAVIDPDYLY
SQINGNALISIPYAKFGMGAEULTKLMEFLFAEEMGGYPRVPIQRARILNE
VQKITHTMLTKTDQDFLKDQITISGKYFQYFFENCQDDEVAAYLKEVLAK"
complement(3089. .3904)
/genes="gno"
/note="spr0290"
complement(3089. .3904)
/genes="gno"
/EC number="1.1.1.69"
/codon_start=1

gene
/transl_table=11
/label="spr0290"
/product="5-keto-D-gluconate 5-reductase"
/protein_id="AAK99094.1"
/db_xref="GI:15457843"
/translation="MTNTSFSIEQSLKGIKALITGASYGIGFAIKSYAEAGATIVF
NDINQDLVNGIEAVREVIGIEAHGVYCDVTDDEGIQAMVKQIQEYGVVILLVYNAGI
IRRVPCMSAADFRKVIDIDLNAPHIVSKAVIPSMIKKGHGKIINICSMWSEIGRET
VYAAAKGGLKMLTRNIASEYGGANICQNGIGPGYIATPQTAPLRELQEDGSRHPFD
QFTIAKTPAARWGNPEDLMGPAVFLASDASNFVNGHILYVDGGILAYIGKQPE"
4175. .4609
/genes="PTS-EII"
/note="spr0291"
4175. .4609
/genes="PTS-EII"
/note="Putative"
/codon_start=1
/transl_table=11
/label="spr0291"
/product="Phosphotransferase system sugar-specific EII
component"
/protein_id="AAK99095.1"
/db_xref="GI:15457844"
/translation="MKIVLVGHGHPATGIYSSLIQAGNQENVEAIDFVEGMSADELK
OKILLAISSNEEVLILSDLLGSGPPKVSSTINGENAKTMVLSGLNLAMLEAVPAR
MAHSPDEVNKSVAQAQGVVNGKELFSDAEEEDFESGI"
4621. .5811
/genes="ugl"
/note="spr0292"
4621. .5811
/genes="ugl"
/codon_start=1
/transl_table=11
/label="spr0292"
/product="Unsaturated glucuronyl hydrolase"
/protein_id="AAK99096.1"
/db_xref="GI:15457845"
/translation="MIKKVITIEKISPERFEVPLLTKEEVQAIDKVIQLEINLDY
FKEDFPPTAFDNYVPIMDNTIEWTGFWTGELWAYESQQDAFKNAHKNVLSFLDR
VNRVELDHDHDLGFLYTPSCMAKYKINGGEAREATKAADKLIERYQKGGFIQAWG
DLGKKEHYRLIIDCLNIQLLPFAVQETGDKYDIAESHFYASANNVIRDDASSPHT
FYFDPETGQPFKGVTRQGYSDSDSCWARGSWGYYGPIVYRHLKDSCTDLPKGVTVY
FANRLPKDHYVMDLIFNDGSDQSDSSATAVAVCGIHEMLKHLPEVDADKDIYKHAM
HAMLASLIEHYANDQFTPGTSLHGVISWHSGKGVDEGNIMGDIYILEALIRFYKDM
NLYW"
5822. .6313
/genes="PTS-EII"
/note="spr0293"
5822. .6313
/genes="PTS-EII"
/note="Putative"
/codon_start=1
/transl_table=11
/label="spr0293"
/product="Phosphotransferase system sugar-specific EII
component"
/protein_id="AAK99097.1"
/db_xref="GI:15457846"
/translation="MTMPNIIMTRIDERLIHQGGQWVKYLCGNTVIVANDEVSTDKM
QQTLMKTVVDSVAMRFFPQKVIDIIHKANPAOTFIIVKDKQALTIVEGVPIKE
INIGNIHAPGKEQVTRSFILGEEDKAALKELSQTHQVTFNTKTPTTGNDAVQVNM
DYI"
6328. .7107
/genes="PTS-EII"
/note="spr0294"
6328. .7107
/genes="PTS-EII"
/note="Putative"
/codon_start=1
/transl_table=11
/label="spr0294"
/product="Phosphotransferase system sugar-specific EII
component"

```

/protein id="AAK99098.1"
/db_xref="GI:15457847"
/translation="MSINVFQAILLGLWTARCFSGMILLGIYVNRGIVLSPFGVILGLD
LPTALAMGALGELAYMGVGAGGTVPNPPIGPIGFIPTLMAITSAKGVSPRAALST
PIAVIQLOTFAITVRAGAPETAKMLNHLKKFPTLNATLWLFAPIGFTIGCLG
ALSDTLKLVVDYIPVLLTGLTVAGKMLPAIGFAMILSVNAKKELIPFVLGYCAA
YLNITIGIAVGTIFALIEFYFNKPKTADHVVEBAHDDWI"
7094..7912
/gene="PTS-EII"
/label="PTS-EII"
/label="spr0295"
/label="Phosphotransferase system sugar-specific EII
component"
/protein id="AAK99099.1"
/db_xref="GI:15457848"
/translation="WTGSKNLTQRDYLKTSLRPFQCONGVSNYQGLGVANVMYPAL
KXHGDEGQFOALENECFYNTNPHELFPITSLHLVLENGRPTKTRSRKMLAG
PLAGIDSLQFLAPLESTIAASFAQGLVVGPIPLFLAMNTILITAKLSGLGYK
LGTVIDKLSQMAITSRIANIIGTVIAGLAATSVKIMVPTFAAGRVKADAKOSIV
SIQGMKDVAPALLPALFTLLVYLIKWKWTYKLVILTVIIGISWUKIIA"
7912..8205
gene

Alignment Scores:
Pred. No.: 2,618-05 Length: 11333
Score: 189.00 Matches: 139
Percent Similarity: 36.21% Mismatches: 100
Best Local Similarity: 21.06% Mismatches: 273
Query Match: 5.41% Indels: 148
DB: 1 Gaps: 28

US-09-802-285A-2 (1-659) x AB008410 (1-11333)
QY 71 SerLysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnPro 90
DB 8302 TCARAAGACTATGAGAAGTCAAGAGTCCCTTGACCGTTGATGACAAATCGTTTATG 8361
QY 91 IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis 110
DB 8362 TTTGATAGT---CCTTGGATATGGAGCCTTGTTCAAAATCCATCAAAATCCAGCG--- 8415
QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
DB 8416 -----ATGGTATGGATCAAGTATTTGAAGAT 8442
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpGlnAlaMetAlaLeuVal 150
DB 8443 GATCCAGATGCTTATATGCTCAATCGACAAGATATCTCTGCAGTTTATGATAGG 8502
QY 151 TyrHisAlaThrGlyAspGluLysTyrAlaArg-----GluTrp 163
DB 8503 TATCTGGTAGAGGAGATAAGGCTATATCCAAAAGTCAAGTCTCTTCTATTGATTGG 8562
QY 164 ValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsn 183
DB 8563 ATTGACGGTGAGAGAAATTTCTCCTCAATCTTGATGACTAGAACCTTGGATACGGT 8622
QY 184 -----LysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProPro 201
DB 8623 ATTCGTTCTTTACTTTGGTTGAACHTA----- 8649
QY 202 ThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeu 221
DB 8650 ---CTCTTGCTCTCTCTTAAATTTGACTTGTGCTAGAGGAGAAAGAACTAGAGAAATTTTG 8706
QY 222 AsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln----- 238
DB 8707 GTCAGTCTAGAAAAGCAGATTGACTTTATGAAGAAGCTACTATCGGCCAAGTACACCCCT 8766

```

```

QY 239 GlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGlu 258
DB 8767 AGTAACCTGGGAGTTTACAAACAATTCGGATGCTGCTATCTATCTATCTTTTCAGAT 8826
QY 259 PheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluLeuLys 278
DB 8827 AAGATGGATCTAGAAAAGCTTACCATTGCTTGCAGAGGAGTTCGAAACAGCAAAATGAG 8886
QY 279 LysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisVal----- 296
DB 8887 ACACAGATTTTAGGAGATGGAAGCAGGTTCGAACAGTCGATCTCTATCATGTAGAGTT 8946
QY 297 -----AlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu 314
DB 8947 TATAAAGCCTTGCTGATTTGTGCTC-----TTG 8976
QY 315 GluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle 334
DB 8977 CTTCCAGACTTTCGAAGATAGTTTACCAAGAGTTGCTGGAAGAGATGCGCACCTATATCAA 9036
QY 335 SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLys 354
DB 9037 ATGATGACAGGCTTAGATGCGGACTTTGGCTTTTGGTAGAGCAATCTACAGAACG 9096
QY 355 AsnPheArgMetAlaGlnPheAla----- 362
DB 9097 ACAGAAATGTTGAGCCTGCTGCTGTTTGAACAAGGAGACCTTCTTAACGCTCTG 9156
QY 363 -----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPhe 377
DB 9157 CATGTTAAAGTTGATTTGCTTAGCATCTTGTTCCTGGGCGAGAAAGTCAAGCGACTG 9216
QY 378 AlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAla 397
DB 9217 CAGGAATTTGAAGGAGAGCTTGGCAGCCTAAG-----TCCATGATCTTTGAAGACTCT 9270
QY 398 GlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAla 417
DB 9271 CGACATGCTGCATTAAGGAT-----GACATCGTTATCTATTTTCAAAAT 9318
QY 418 SerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluPheIleLys 437
DB 9319 GGTCCGCTAGGAAGTGCCTATAGCCATAGTACGAGATAAGTTTGTCTACGATATCAA 9378
QY 438 GlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMet 457
DB 9379 GCCAACCTATTTTATAGATGCTGGCGTTATCTTAT----- 9417
QY 458 LysLeuArgAsnTrpTyrArgGlnThrArgIle-----HisSerThrLeu 472
DB 9418 -----CGGAGATATATGAACGTTATCTCTTAAAGAGTCTTGGAGTCAATTCGACCTGC 9471
QY 473 ThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGln-----AsnLysTrp 489
DB 9472 ATTGTAGATGGG-----AAAGCTCCGAAAGAAATACGGGATCCTCG 9513
QY 490 Glu-----ThrGlyAsnAsnLeuAspValLeuThr 499
DB 9514 GAATATGATATATCTCCTCACTCCCTGTTTGTCCACCATAAAGAAAGGAGGAATGCAT 9573
QY 500 TyrThrAsnProSerTyr-----ProAsnLeuAsp-----HisGlnArgSer 513
DB 9574 TATATTGAAGGGGCTTATTTGGTCAGCAGAACCTGATTTGCCCTATCTTCAACAGAGAAA 9633
QY 514 ValLeuPheIleAsnLysLysTyrPheLeuValIleAsp-----ArgAlaIleGlyGlu 531
DB 9634 ATCCCTCATGTTGGTAGAGGATCTCGCTCTTGGTAGATGACATCAGGTCTCAAGTCTAG 9693
QY 532 AlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAsp 551
DB 9694 -----CATGAGCGGTGACTCAGTTT-----ATCCTTGAC 9723
QY 552 LysThrLysAsnArgValTyrThrTyrArgAspGly-----AsnAsnLeuMetIle 569

```



```

gene
/IDMLIGLVIIRAADALIE"
/complement(2577..3311)
/gene="SCF62.04c"
/note="SCF62.04c"
CDS
/complement(2577..3311)
/gene="SCF62.04c"
/note="SCF62.04c, possible secreted protein, len: 244 aa.
Highly similar to Streptomyces coelicolor TR:CA852958
(EMBL; AL109950) hypothetical 25.2 KD protein SC74.24C
(242 aa), fasta scores opt: 1283 z-score: 1457.6 E(): 0
83.1% identity in 242 aa overlap. Contains a possible
N-terminal signal sequence."
/codon_start=1
/transl_table=11
/product="putative conserved secreted protein"
/protein_id="CA858319.1"
/db_xref="GI:6066659"
/db_xref="SPTREMBL:Q9RJL0"
/gtranslation="VRMKIAVIGGTGLIGSQVVKDNLNAGHQAIPHRSRSTGVDVVGGG
GVDDAVEAGAEVVNLTNSPTFDDASPAFFRASMDTLLAAARRAGTGHFVILSVGVDR
VPALDYIRAKVLOEILRDGPVPYIVRATQPMERMEAVLSWTASDDSVRLPATPIOP
TAAKDVAARAVADVAGPLNGIRINIGGPEVFLDELGLRLTLAKHGDAITVYVDPAGM
FGAVEGDLTDLDAHAPTRYADWLS"
/complement(3320..3324)
3580..3584
RBS
3591..5054
RBS
/gene="SCO0379"
/note="Kata"
CDS
3591..5054
/gene="SCO0379"
/note="SCF62.05, kata, catalase (EC 1.11.1.6) len: 487 aa.
Has been previously sequenced and characterised from
Streptomyces coelicolor strain ATCC10147 TR:P77948 (EMBL;
X96981) catalase (EC 1.11.1.6) (488 aa), fasta scores opt:
3300 z-score: 3768.6 E():0.99.2% identity in 488 aa
overlap. Also similar to another proposed Streptomyces
coelicolor catalase (EC 1.11.1.6) TR:Q9Z598 (EMBL;
AL035478) SC2G5.25C (487 aa), fasta scores opt: 1704
z-score: 1896.1 E():0.53.8% identity in 483 aa overlap.
Contains Prosit hits to PS00437 Catalase proximal
heme-ligand signature and to PS00438 Catalase proximal
active site signature. Also contains a Pfam match to entry
PF00199 catalase, Catalase."
/codon_start=1
/transl_table=11
/product="catalase (EC 1.11.1.6)"
/protein_id="CA858320.1"
/db_xref="GI:6066660"
/db_xref="SPTREMBL:Q9RJK9"
/gtranslation="MPENNQKPLTTVAGAPVPDNONSLTSGRPGMLLDQVWFLEKLA
HFDRVIERMRHAKSGAGFGFTVTHTDITRYTSAKIFSEIGKKTPLFRSTVAGER
GAADAEIRGPFAVKFYTDGNDWLVGNNTPVFFRDLKPELDNHAHKRPTNLRN
AENWDFWTLNPEALHQTIVMSDRGIPASVYHMGFGSHYSLINABGRFWFKFHH
RIQQIKNLTDAAEALVKQRESHQDRLFDAIEDGDFPKWLFTQVMPEDAENYRF
HPFDLTWKKXKDYPLIRVGEWELNPNPNFADYQEAFAFSPANTVPGISFSPDRLQ
GRLFYSQDAQYRLGVNHHQIPVNAFKNPVPNSYHRDGMRVGDNQCATPGVEPNYSYGR
WQEQPVRDPAQAVADRENYREDDDNYEPQGNLFRQMSPEQQLFEENTARAID
GASATPIERHIGNCTQADPAYGAGYRKALAEALAGNL"
3618..5030
/misc_feature
/gene="SCO0379"
/note="Pfam match to entry PF00199 catalase, Catalase,
score 1129.70, E-value 0"
3726..3776
/misc_feature
/gene="SCO0379"
/note="PS00438 Catalase proximal active site signature"
4596..4622
/misc_feature
/gene="SCO0379"
/note="PS00437 Catalase proximal heme-ligand signature"
5090..5132
/stem_loop
/note="Inverted repeat 21/21 (100%) with 1 base loop."
5145..5148
RBS
5158..5553
gene
/gene="SCF62.06"

```

```

CDS
/note="SCO0380"
5158..5553
/gene="SCF62.06"
/note="SCF62.06, unknown, len: 131 aa. Weakly similar to
several prokaryotic Ank-like proteins e.g. Pseudomonas
syringae TR:P95546 (EMBL; AF001355) AnkF precursor (183
Alignment Scores:
Pred. No.: 0.000185 Length: 34496
Score: 186.50 Matches: 105
Percent Similarity: 33.33% Conservative: 61
Best Local Similarity: 21.08% Mismatches: 198
Query Match: 5.34% Indels: 135
DB: 1 Gaps: 18
US-09-802-285A-2 (1-659) x SCF62 (1-34496)
Qy 58 AspAlaAlaLysAlaLeuLeuAla-----tyr 66
Db 17535 GAGCGCGGAGAGCGCTCGTCCCGGAGCGGACCGCTGCTGCGCGGACGTCGAGTAC 17594
Qy 67 TyrArg---GluLysSerLysAlaArgLnuProAspPheSerAsnAlaGluLysProAla 85
Db 17595 TTCGGCGTGGCGCGGACGACCTGCTCGAGCGGACTGGTGC-----17636
Qy 86 AspIleArgGlnProIleAspLysValThrArgGluMetAlaAspLysAlaLeuValHis 105
Db 17636-----17636
Qy 106 GlnPheGlnProHisLysGly-----TyrGlyTyr---PheAspTyrGlyLys 120
Db 17637---TTCGACCCGAGACGCGGCGCGCGCGCGCGCTGCGGCTACGCTTCGACGTG---17687
Qy 121 AspIleAsnTrpGlnMetTrpProValLysAspAsnGluVal-----134
Db 17688-----CCGTACCCGACGAGGACGCGGTCGGTGCATCAAGCAG 17726
Qy 135 ArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThr 154
Db 17727 ATCTGGGAGCGTCCCGGACACGACGACCTACCGCTGCTGCGCGGCGCTACGCGGTAC 17786
Qy 155 GlyAspGluLysTyrAla-----ArgGluTrpValTyrGlnTyr 167
Db 17787 GCGCAGCAGCGGTACGCGCGCGTGGCCGAGCACCTGCGTCTGTGG-----17834
Qy 168 SerAspTrpAlaArgLysAsnProLeu-----GlyLeuSerGln 180
Db 17835-----TGGCGCGCCCAACGCGCGCTGCGCGGGTGCACCTGGGTGACGCGCATCGAGCTG 17888
Qy 181 AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro 200
Db 17889 GGGATCCCGCTGCTCTCTCTG-----GTGTGGATCCCGCGGCTGCTGACGGCTGGCGG 17942
Qy 201 ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220
Db 17943 GCGCGCGCGCGCTCTTCGAGGACAAACCGCGG-----CGCGTGGAGCAG 17987
Qy 221 LeuAsnSerTyrHisGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsn 240
Db 17988 ATCTGGCACCCACGAGCGTGGCTGGCGCGCTTCCCGACGCGGGTCTGTCGCGCAACAC 18047
Qy 241 HisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys 260
Db 18048 CACGTCTGGCGGAGCGCGCGGCGAGTTGGCGCGGCGCTGCGCTTCGGGTGGTTCGCC 18107
Qy 261 AspSerProArgTrpArgGlnThrGlyLysSerValLeuAsnThrGluLysLysGln 280
Db 18108 GAGTCGCGCAGCGTGGCGGACGACGCGCTGCGGTGCGTGGAGCGGACCTGCGCGGCAAC 18167
Qy 281 ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaLeuAsp 300
Db 18168 AGTTTCTCTCGCGCTCAACCGGAGCTGGCCACCGAGTACACCGGCTGGTGTGGAA 18227

```

```
QY 301 IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln 320
Db 18228 CTCGGCTGGCGGCTGGCGGAGGCGGAGCGCGCGGTGTG-----CGGTGCCCGCG 18281
QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerLeuProAsp 340
Db 18282 ACGTCCGGCTGGTGTGGCGATGACGACGCGCTCGCGCCGCTCGTGGACGACCGG 18341
QY 341 TyrAsnThrProMetPheGlyAspSer----- 349
Db 18342 CTGCGTCCCGCGCGGAGCGGACGCGACGCGCGCTGGTGGACGCGCGG 18401
QY 349 ----- 349
Db 18402 GGCACCGACCGCTGGGCTCGTGTCTGCCACCGGGGACCGGTTCGGTGGCTTCCC 18461
QY 350 ---Tip-----IleThrAspLysAsnPheArgMetAlaGlnPheAlaSerTrpAlaArg 366
Db 18462 TGGTGGCGGACGCGTACCGCGGATGTGCGCACCGCGTGTGGCGGCTGGTGGTCCGG 18521
QY 367 ValPheProAlaAsnGlnAlaIleTyrPheAlaThrAspGlyLysGlnGlyLysAla 386
Db 18522 CCTACGCGGCGGACGACG-----GGCGGCGCGTGCACCGCG 18560
QY 387 ProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyrThrPheArgSer---Gly 405
Db 18561 CGCGCGCGGACGCGCGCCACTTCGCGGACGCGGCGCTACGGTGTGCGGCGCGCGG 18620
QY 406 -TrpAspLysAsnAlaThrValMetValLeuLysAlaSerProGlyGluPheHisAl 425
Db 18621 GGGATCTGGTGGCTGCGACGCGGCTCGCGACGCGCTTCCTGTCATCGCGCG-CACGC 18679
QY 425 aglnProAsnAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAl 445
Db 18680 CCACGCGGACGCGCTCTCGTGGAGTCCGCGACGCGGCGTGCAGCTCGCGACCC 18739
QY 445 aglyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGl 465
Db 18740 GGGGACCTACTCTACCGACGACGCGGTG-----TGGCGGGGTACTTCGGTTC 18790
QY 465 nThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThr 482
Db 18791 GACGCTCGGCCCAACACCTCGAGTGGACGCGCGCGACCGTCCGCTCC 18842

RESULT 9
AF498407 22913 bp DNA linear BCT 13-JUN-2002
LOCUS Pseudomonas aeruginosa serotype 015 putative O-antigen biosynthesis
DEFINITION gene cluster, partial sequence.
ACCESSION AF498407 AC104726
VERSION AF498407.1 GI:20559869
KEYWORDS
SOURCE
ORGANISM Pseudomonas aeruginosa.
Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 22913)
Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kutyavain,T.V.,
Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.
Genetic variation at the O-antigen biosynthetic locus in
Pseudomonas aeruginosa
J. Bacteriol. 184 (13), 3614-3622 (2002)
JOURNAL 22053227
MEDLINE 12057956
PUBMED 12057956
REFERENCE
2 (bases 1 to 22913)
Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kutyavain,T.V.,
Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.
Direct Submission
TITLE
Submitted (04-APR-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98105-2145, USA
JOURNAL
COMMENT On May 13, 2002 this sequence version replaced gi:17975255.
FEATURES
Location/Qualifiers
1. .22913
source
```

```
CDS
/organism="Pseudomonas aeruginosa"
/serotype="015"
/db_xref="taxon:287"
<1..104
/feature="ORF 1; C-terminal coding region of rpsa"
/codon_start=3
/transl_table=11
/product="rpsa"
/protein_id="AAM27639.1"
/db_xref="GI:20559870"
/translation="EKDAMKELRKQEVESAGPTTIGDLIRAQWENQG"
241..525
/feature="ORF 2; himd/iHFB"
/codon_start=1
/transl_table=11
/product="Himd"
/protein_id="AAM27640.1"
/db_xref="GI:20559871"
/translation="MTKSELIERIVTHOGQLSAKDVELAIKTMLSQMSOALATGDRIE
IRGSGSFLHYRAPVGRNPKTGESVRLDGKEVPHFKGKEDRDRVNEPE"
359..861
/feature="ORF 3; hypothetical orfa"
/codon_start=1
/transl_table=11
/protein_id="AAM27641.1"
/db_xref="GI:20559872"
/translation="MOWFRSAVAVALLVSTAVVVFVLENGVVLQALFLGVQFPQWP
LAVYIIAIFILGGLGLAAQFPFLTISRVRASSLRAELVQARKEVDTLRAEPSKVS"
1146..2174
/feature="ORF 4; wzz; similar to chain length determinant
protein"
/codon_start=1
/transl_table=11
/product="Wzz"
/protein_id="AAM27642.1"
/db_xref="GI:20559873"
/translation="MSKNEIDFIELIOGLWAQKVLVSLTFLGVALAGSVAFLSKPV
YEARVAVPSSLSDVAGNIARSDSGLPFTVSQVYSVFTNLSGTRQFFREVY
LPSLEERVGDLVAFSDVSNQPTKELPERYTVVFDHPDPPQAKWAKQYIELA
AQRSDMERGNAQRELVSRNIQQOITLQVVKERRDTRIVRKEALQVAEAVGVE
KPLATGSQIDQOLSIMEGSLMYMRGSRALRAEIALBARTSDDPFIPALRGLQOLA
LVESKQLSFDRVAVPQDGSVETPDDPIKPRKVLIALGILAGGVLGVLAALLRLFR
EKELDTG"
2421..3491
/pseudo
/codon_start=1
/transl_table=11
3524..4786
/feature="ORF 6; similar to UDP-glucose/GDP-mannose
dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="AAM27643.1"
/db_xref="GI:20559874"
/translation="MSFOTISVIGIGYIGLPTAAVFASRQKVKIVGVNQAAVDITNR
GEIHVPEPDLMTVHAATGTYLRTATNPEDAPADFLIAVTPFFKHEHSDIRVIBAS
KSIAPVLKKGDLVILESTPGVATEQMGALAEFPDLTFPTQHEASDIRVIFBAAS
VLPGHRELIVENDRVICGSEKSEAAARLYKIFVGEICITINARTAMCKLTNSF
RDVIAFANELSIIICDLIDINWELIRLANRHPRVNILOPGVGCGHCTAVDPWFTVS
KTPDARLIRSAREVNDHKPFWLEKIKLAVADVLOKPEKTARDVSVACFLAFKAN
IDDLRESPAJIAQOVAVASHPGRVLVPEPNIIEELPKFSGNLTLASTDSMQEADVIV
LLVHKFEPEIAPHPRQVVDITRGLWEA"
5040..6155
/feature="ORF 7; similar to Polysaccharide biosynthesis
protein; potential multiple membrane spanning domains."
/codon_start=1
/transl_table=11
/protein_id="AAM27644.1"
/db_xref="GI:20559875"
/translation="MLIHALVGYASIFDAGLGKSVVROIAISSAPEKRRGPTGLSSLL
FVLISGLAALVHFVFAEYLVVDIPKVDSSYEDALTGTVAALCIPFLASLVLOQY
LEGVEDFYSFNKRALSGTLTYLLPVAFLLPEKTFSSLIYGLFARAGSCFLALIVS
RCSVWTRWMDVKVFKTLISFGSWLITNIISPVNVMYMDKFLIARITAGELAYTAA
```



```
QY 275 ThrGluLeuIleLysLysGlnValTyrAlaAspGlyMetClnPheGluLeuSerProIleTyr 294
Db 13572 AACCGGGCGNAACATCTGATTCGGAGGAGGAGTTCACCGAGTATTCGCTGAATTAC 13631
QY 295 HisValAlaAlaIleAspIlePheLeuLysAlaIleTyrGlySerAlaLysArgValAsnLeu 314
Db 13632 CACCGGGTCATGCTGATACGTTCTCTACTGCGGAGCTGGAGGCGTCAGCTTCCCTT 13691
QY 315 GluLysGluPheProGlnSerTyrValGlnThr-ValGluAsnMetIleMetAlaLeuI 334
Db 13692 GTGGAATTCCTGCTGCTGATTCGCGGCAAGTCGACGACTCACTGCTGCACGCA 13751
QY 334 eSerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSer----- 349
Db 13752 TTCATTCAGCCTGCCACT-GGTACGGTCCGAACCTTTGGGCTTAACGATGGTCACGACT 13810
QY 350 ----TriIleThrAspLysAsnPheArg----- 359
Db 13811 GCTACCTCTGACCCAGACCGATTCATCGTACTACGACCTGCGTGCAGTTCGCCATGGC 13870
QY 359 aGlnPheAlaSerTyrAlaArgValPheProAlaAsnGln-----AlaIleLys 375
Db 13871 ATGTTCACCGAT---ACCAACGCCCTATCCCGCAATCAAGAGTCGACGCCGCGCTGCT 13927
QY 375 sTyrPheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSe 395
Db 13928 GTGGCTGGGCTCGATCGCTCTGCGGAGTCTCTGTACCCGAGCGCGAGTGTTCGA 13987
QY 395 rAsnAlaGlyPheTyrThrPheArgSerGlyTyrAspLysAsnAlaThrValMetValle 415
Db 13988 CGAGGGTGGCTACGCCATTCTCGGAGCGG-----CAGCCAATGCCAAT 14032
QY 415 uLysAlaSerProGlyGluPheHis----AlaGlnProAspAsnGlyThrPheGluLe 434
Db 14033 GCTGCGTTACCCAGGTTCCGTTTTCGCCCGCAGCCAGCGGCGTCTTCATCTGATCT 14092
QY 434 uPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspG 454
Db 14093 TTGGCTCGAAGGGCGAAATCTGTCGCGCATGCGGCGCCTACAGCTACACACCGCAATC 14152
QY 454 uAlaIleMetLysLeuArgAsnTyrPyrArgGlnThrArgIleHisSerThrLeuThrLe 474
Db 14153 CGCTTGGCTG-----AGTTACTTCCCGGACGCCAGTCACACAGATCCAGTT 14203
QY 474 uAspAsnGlnAsnMetVal-----lleThrLysAlaArgGlnAsnLysTyrPheGlu 492
Db 14204 CGATGTCGCGATCAGATGCCCGCTCTCAGCCGTTTCTTTTGGCGACTGTTGAAGGC 14263
QY 492 yAsnAsnLeuAspValLeuThr----- 499
Db 14264 GACCGAGGCTGTCGCTGACTCGGAGCGAGCTCGGAGAAATGAATGATTTTCGGC 14323
QY 500 ----TyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerVal---LeuPheI 517
Db 14324 GGCCTATACCGAT---GGGTAT---GGCGCAGGATAGCGGAGGTTCCCTTGTTCC 14377
QY 517 eAsnLysLysTyrPheLeuValIleAspArgAlaIleGlyGluAlaThr----- 533
Db 14378 GGATGGA-----CTCAAGGTTCATGATGAGATTTCCGGCTTTTCGCGAGTCCGCTGCT 14431
QY 534 -----GlyAsnLeuGlyValHisTrp-----GlnLeuLysG 544
Db 14432 GCCTCGAGGCTGGACCTGGAG-CTGGTGTGTGGATGAAACACGCTCAAGGGAGGA 14490
QY 544 uAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrTyrArgAspG 564
Db 14491 GGGGAGAGTGTGCTGATTTTCGAAGCTCGAT-----GCCTCTTCGCGTATCGACTGA 14544
QY 564 yAsnAsnLeuMetIleClnSerLeuAsnAlaAspArg----- 576
Db 14545 CAATCGGGTGGAGTGGCGCTACTACTCTGCGAAGAGGAGGTTCTCTGTCATCGAGTAG 14604
```

```
QY 577 -----ThrSerLeuAsnGluGluGly-----LysValSerTy 588
Db 14605 AGTTGGGCGAGCGGGAACATTCAGACGGATTACAGTGGGCCCCGATGAGATCATCTA 14664
QY 588 rValTyrAsnLysGluLeuLysArgProAlaPheValPheGluLysProLysLysAsnAl 608
Db 14665 TTTTCATCAACAC-----TTTTCAGTCCGAAGGATGGT 14700
QY 608 aGlyThrGlnAsnPhe 613
Db 14701 AGGCATCCGCTCCTAC 14716

RESULT 10
BD003866/c
LOCUS BD003866 3763 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003866
VERSION BD003866.1 GI:18631827
KEYWORDS JP 2001501833-A/186.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3763)
AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C.,
Ramon,M. and Dougherty,B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 186 13-FEB-2001;
HUMAN GENOME SCIENCES INC
COMMENT OS Unidentified
PN JP 2001501833-A/186
PD 13-FEB-2001
PF 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
STEVEN C BARASH,
PI MICHAEL FANNON, BRIAN A DOUGHERTY
PC C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19,
PC C12N1/21,
PC C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC
G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source
FT 1..3763 /organism='Unidentified'
BASE COUNT 1178 a 804 c 677 g 1104 t
ORIGIN

Alignment Scores:
Pred. No.: 7,21e-05 Length: 3763
Score: 176.00 Matches: 116
Percent Similarity: 34.63% Conservatives: 80
Best Local Similarity: 20.49% Mismatches: 231
Query Match: 5.04% Indels: 140
DB: Gaps: 20

US-09-802-285A-2 (1-659) x BD003866 (1-3763)

QY 71 SerLysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnPro 90
Db 2173 TCAAAAGACTATGACAGAAAGTCAAGAGTCCCTTGAACGCTTGATGGACATCGTTTATG 2114
QY 91 IleAspLysValThrArgLysAlaLeuValHisGlnPheGlnProHis 110
Db 2113 TTGTAGT---CCTGGGATGAGGCTGTTCAAAATCCATCAATCCAGCG--- 2060
QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
Db 2059 -----ATGGTATGGGATCAAGTATTTGAAGAT 2033
```


Qy	131	AspAsnGluValArgTyrGlnLeuHisArgValLysTyrTrpGlnAlaMetAlaLeuVal	150	Qy	458	LysLeuArgAsnTyrTrpArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGln	477
Db	2032	GAICAGAAATGCTTATATGCTCAATCGACGAAGATATCTCTTCAGATTATGATAGG	1973	Db	1037	TCTCTAAAGAGTGTGGAGTCATTCGACCTGCATTGT	999
Qy	151	TyrHisAlaThrGlyAspGluLysTyrAlaArg	163	Qy	478	AsnMetValIleThrLysAlaArgGlnAsnLysTyrGlu	496
Db	1972	TATCTGGTAGAAGAGATAAGGACTATATTCAAAGTCAAGTCTCTTCTATTTGATTGG	1913	Db	998	AGATGGGAAAGCTCCGGAAGAATCACGGG	969
Qy	164	ValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsn	183	Qy	497	ValLeu-ThrTyr	500
Db	1912	ATTGACGAGGTGAGAGAATTTCTCTCAATCTTGATGACTAGAACCTTGATGATGGGT	1853	Db	968	ATCCTTGGGAATATGAATACTATCTCCTCCTGTTTGTTCACCAATAAGAAAGGAGGG	909
Qy	184	LysPheValTyrArgProLeuGluValSerAspArgValGlnSerLeuProPro	201	Qy	501	ThrAsnProSerTyrProAsnLeuAspHi	510
Db	1852	ATTGCTTCTTTACTTGGTGAACATA	1826	Db	908	AGTGCATTATATTAGGGGGCTTATTTGGTCACAGAACCTGATTTCCTTATCTT	852
Qy	202	ThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeu	221	Qy	510	scLlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAsp	528
Db	1825	CTCTTGCTCTCTTGAATTTGACTTTGCTAGAGAGAGAAAGACTAGAGAAATTTTG	1769	Db	851	CAGGAGAAAAATCCTCATGTTGGTAGAGGATCTCTGGCTCTTGTGTAGATGACATCAGGTG	792
Qy	222	AsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln	238	Qy	528	atleGlyGluAlaThrGlyAsnLeuGlyValHisTyrGlnLeuLysGluAspSerAsnPr	548
Db	1768	GTCAGTCTAGAAAAGCAGATTGACTTTATGAAAAGCTACTATCGCGCCCAAGTACACCCCT	1709	Db	791	TCAAGGTCAA	761
Qy	239	GlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGlu	258	Qy	548	oValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGly	566
Db	1708	AGTAACCTGGGATTTTACAAATCCGATCGCTGTCTATCTATCATTTCTTTTCAGAT	1649	Db	760	ATCCTTGACAGGAT	717
Qy	259	PheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluLeuLys	278	Qy	566	nLeuMetIleGlnSer	571
Db	1648	AAGATGGATCTAGAGAAGCTTACCATTGCTTCAGAGAGTGTGAACACAGAAATTTGAG	1589	Db	716	GTTCAGACTATGGAGT	701
Qy	279	LysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisVal	296	RESULT 11			
Db	1588	ACACAGATTTTAGGAGATGGAGCCAGTTTGAACAGTCTCTATCTCATCTAGAGGTT	1529	AE007344	14385 bp	DNA linear	BCT 31-AUG-2001
Qy	297	AlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu	314	LOCUS			
Db	1528	TATAAAGCTTCTGATTTGTCTC	1499	DEFINITION			
Qy	315	GluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle	334	VERSION			
Db	1498	CTTCACAGCTTGCACAGATTTACCAAGAGTTGCTGGAAGATGCGGACCTATATCAA	1439	KEYWORDS			
Qy	335	SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLys	354	SOURCE			
Db	1438	ATGATGACAGGCTTAGATGGACGGACTTTGGCTTTTGGTGATAGCGATTCTACAGAAACG	1379	ORGANISM			
Qy	355	AsnPheArgMetAlaGlnPheAla	362	REFERENCE			
Db	1378	ACAGAAATTTTGAGCTGCTGCTGTTTGAACAGAGACCTTCTTAACGGTCTG	1319	AUTHORS			
Qy	363	SerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPhe	377				
Db	1318	GATGTTAAAGTTGATTTGCTTAGCTCTTGTCTGGGGCGAGAAAGCTCAAGCAGCTG	1259	TITLE			
Qy	378	AlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAla	397	Complete genome sequence of a virulent isolate of Streptococcus pneumoniae			
Db	1258	CAGGAATTTGAAGAGAGCTTGGCAGCCTAAG-----TCCATGATCTTTGAAGACTCT	1205	JOURNAL			
Qy	398	GlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAla	417	SCIENCE			
Db	1204	GGACATGCTGCATTAAGGAT-----GAACATCGTTATCTATTTTCAAAAT	1157	MEDLINE			
Qy	418	SerProProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheLys	437	PUBMED			
Db	1156	GTTCGCGCTAGGAAGTCCCATAGTCTAGTACGAGAATAGTTTTTGTCTACATATCAA	1097	REFERENCE			
Qy	438	GlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMet	457	AUTHORS			
Db	1096	GGCCACCTATTTTCATAGATGCTGGCGTTATCTTATCG-GGAGATATATGAACGTTA	1038				

Streptococcus pneumoniae TIGR4 section 27 of 194 of the complete genome.

Streptococcus pneumoniae TIGR4.

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

Streptococcus.

1 (bases 1 to 14385)

Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,

Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,

Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,

Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D.,

Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L.,

McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T.,

Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O.,

Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and

Fraser,C.M.

Pneumoniae

Science 293 (5529), 498-506 (2001)

21357209

11463916

2 (bases 1 to 14385)

Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,

Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,

Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,

Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E.,

Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A.,

Feldblyum,T.V., Angiuoli,S., Gesuwan,P., Hickey,E.K., Holt,I.E.,

Loftus,B.J., Ujwal,M.L., Yang,F., Smith,H.O., Venter,J.C.,

Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.

Direct Submission

Submitted (29-JUN-2001) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1.14385

source


```
/organism="Streptococcus pneumoniae TIGR4"
/db_xref="TIGR4"
complement(142..618)
/gene="SP0313"
complement(142..618)
/gene="SP0313"
/notes="identified by match to PFAM protein family HMM
PF00255"
/codon_start=1
/transl_table=11
/product="glutathione peroxidase"
/protein_id="AAK74490.1"
/db_xref="GI:14971787"
/translation="WTSLLDFSVLNNQNAOPLDSYRGKVLIVNTATCGLTPTQVOG
LELRYVQDQFEILDPCFMQAGSABEINAFCSLHFQTFPRFAKIVNGKE
ADPLVYMQDQSGPLGRVFNFAKFLIGRDGVFERFSSKDPQIEAIIQTLL"
841..4041
/gene="SP0314"
841..4041
/gene="SP0314"
/notes="identified by match to PFAM protein family HMM
PF02278"
/codon_start=1
/transl_table=11
/product="hyaluronidase"
/protein_id="AAK74491.1"
/db_xref="GI:14971788"
/translation="MQTKKLIVSLSSVLGSLNHYMTGABETTNTTIQSQOKE
VVOYQDRTKLVNEDSGTDEGSPWTSKAGWSAWDQKNSADASTRVIEAKDGA
TISSEKLAALARMVPEAKKKYKLRFKITDKNGIKAKVRIIEESCKDLKNSA
TTSQTKDMQTIADYSPTLDDVKIKLEFVETGCTVSEPKDIELVEADOLSEDSQTD
KQLEBKIDLPQKXVFLADYTYKVNFEDVASVNGILPEKGTNNIVSKGKEV
KKIPKILASVDAYTRDDWNGIAGQYVDSKNEQWAKNLEKGVADSLSSIS
SQADRTYWEKESNYKTSANLTATVKLEMAKQVTPSSRYQDETIVRTVSDSMEW
MEKHVYNSEKTVGNWMDYEIGTPRAINNTLSIMKEYFDEEIKKYTDVIEKVPDPE
HFKRTDNEPKALGNLVMGRVKVIAGLRDKDOEISSTIRSEOVFLVQGGFY
QDGSYIDHTNAVYTAGYGNVLDGLSOLLPVQTKNPDKDKMQTMVHIDKSPAL
LVNGLMDMSGRSISRANSEGHVAEVLRIHRIADMSGETKQCSLQSLKTVQS
DSYDVFKNTYKDIOLSMQSLSDAGVASVPSYLSAFNMKDMTAMNAEKGFGF
LSLFSRTLYNHMKENKRGWYTDGMPYLYNGDLSHYSDGYWPTNPKMPTGTET
DAKRAEDSTGVPALFVGTSKLDDANATMDFTNWTNOLTAHKSWMFKDKIAPLG
SNIQNTSDTAATTIDQRLSEGNPVYVNDKEASLTQEKDYPTQSVLESDPSK
KNIGYFFPKSISMSKALQKAWKIDNEGQSDKEVESELATISOAHKONRSDGYML
INVRATPNQMKELLESSLIENNETLQSVYDAKQGWGIKVIDSVSTISNQFVLK
RGVYIRKGGDEYKATYINPETOESAPDOEVFKLEQAAPQVQNSKEKEKESEBKNN
SQQKMLPQKGQCSILASLGILLGAFYLFRRGKN"
4425..5045
/gene="SP0315"
/notes="similar to SP:Pl9780 GB:Y00434; identified by
sequence similarity; putative"
complement(5058..5165)
/gene="SP0316"
complement(5058..5165)
/gene="SP0316"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK74492.1"
/db_xref="GI:14971789"
/translation="WVYLRALSPNHQPAKDGAFHVQVQALLLTRLPL"
complement(5189..5818)
/gene="SP0317"
complement(5189..5818)
/gene="SP0317"
/notes="identified by match to TIGR protein family HMM
TIGR01182"
/codon_start=1
/transl_table=11
/product="4-hydroxy-2-oxoglutarate
aldolase/2-deydro-3-deoxyphosphoglucanate aldolase"
/protein_id="AAK74493.1"
```

```
/db_xref="GI:14971790"
/translation="MTKSDTIIEELKQKIVAVIRGNTKEEGLOASTACIACIKGKAIEI
AYNQYAGQIIKELVDLQDDQSCVICAGTVLDVATARDAILAGANYVSPSPHASTA
KMCNYSTPYIPGCITLITALEAGSEIIKLPFGSTLSPAYISAVKAPIQVSVMV
TGGVGLNNIPQWFAAGADAVGIGGELKSLASQGNFDRISEIAQQYITLR"
complement(5828..6829)
/gene="SP0318"
complement(5828..6829)
/gene="SP0318"
/codon_start=1
/transl_table=11
/product="carbohydrate kinase, pfkB family"
/protein_id="AAK74494.1"
/db_xref="GI:14971791"
/translation="WTKILLPGEPLIRISPLDATSIGDHVASTVFGGSEINIACNLQ
ALGISTKVTALPANEIGDRFLTIKQHQIDTSSICRLGDRIGLYLENGFCGRQSEV
FYDRKHTTISQIRPNMLDMDSLFQGISHFSGITVAIGQEVRAILLILLEAKRGGI
VYSMDLNRITKMSVLEAKYFSPKFTDTCFGLDPLMDQNLMPFDRDASLSEEV
ENRMILLKAYGFKALFHTLRSSDRQDNVQYALAELELFESVOLKTVAYORIGSGD
AFISGALYQLLHSSILKTTIDPAVASATLKTLPDGHLSSTSTSENLLANAQDIIR"
complement(6860..7501)
/gene="SP0319"
complement(6860..7501)
/gene="SP0319"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAK74495.1"
/db_xref="GI:14971792"
/translation="MKIALINENSOAKNHIYDLSKEATDKKQYLFNYGMRGEGE
SOLTVQNGLMAAILNTKAVDFVTGCTGVGAMALNSFGVVGGLAVDPTDAVLY
SOINGNALISIPYAKGFGWGAELTLKMFERLPAEMGGGYPRERVIPQRNARILNE
VQIITHNDLTKIILIDQFLDKTISGTYFQYFFENCODDEVAAYLKEVLAK"
complement(7520..8335)
/gene="SP0320"
complement(7520..8335)
/gene="SP0320"
/codon_start=1
/transl_table=11
/product="oxidoreductase, short chain
dehydrogenase/reductase family"
/protein_id="AAK74496.1"
/db_xref="GI:14971793"
/translation="MNTSFSIEQSLKGLKIALITGASVGIGFAITAKSVAEAGATIVF
NDINQDLNKGIEAREVGIQAHGVCDVDEDDGQAMVKQIEQEVGVDIILVNNAGI
IRRVPMCSAADFRKVIDIDLNAPFIVSKAVIPSMIKKGHKIINICSMGSELGRET
VSAYAAAGGLKMLTRNTIASEYGGANIQCNGIPGVIATPQTAPLRELOQDSRHPFD
QPIIAKTPAARWGNTEDLMPAVFLASDASNFVNGHILYVDGGIILAYIGKPE"
8606..9040
/gene="SP0321"
8606..9040
/gene="SP0321"
/codon_start=1
/transl_table=11
/product="PTS system, IIA component"
/protein_id="AAK74497.1"
/db_xref="GI:14971794"
/translation="MKVILVGHGFATGYSLSQLIAGNQENVEAIDFVEGMSADELK
QKILLAINEEVELLISDLGGSPFKVSTINGENPAKTMVNLGSLNLAAMEAVFAR
NAHSFDEVVKNKSVAAQGGVNGKELFSTDAEEEDFESGI"
9052..10242
/gene="SP0322"
9052..10242
/gene="SP0322"
/notes="similar to GP:5869507; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glucuronyl hydrolase"
/protein_id="AAK74498.1"
/db_xref="GI:14971795"
/translation="MIKVTIEKIKSPERFLEVPLLTKEEVGOAIDKVIROLENLIDY
FKEDFTPATFDNVYPIMDNTIEWNGFTWELWLALEYISQDQAPKNAHKNVLSFLDR
```

VNKRVELDHDHDLGFLYTPSCMAEYKINGDEARBLAKAADKLIERIOYKGGFIQAWG
DLGKKEHVRLLIDCLLNQLLFFAVOETGDKYDYDIAESHEFVASANNVIRDDASSFHT
FYFDPGPKGTQROYSDDSCWARGOSGVYGLPLTYRHLKDESCFDLFGKVINY
FINRFPKHVSYWDLIFNDGSDQSDSATAIACVGIHEMLKHLUPEDVADKDIYKHAM
HAMLSRLIEHYANDQFTPGTSLHGVYSHWSHSGVDEGNIWGDYVYILEALIRFYKOW
NLYW"

gene 10253..10744

/gene="SP0323"

10253..10744

/gene="SP0323"

/codon_start=1

/transl_table=11

/product="PTS system, IIB component"

/protein_id="AAK74499.1"

/db_xref="GI:14971796"

Alignment Scores:

Pred. No.:	0.000429	Length:	14385
Score:	176.00	Matches:	116
Percent Similarity:	34.63%	Conservative:	80
Best Local Similarity:	20.49%	Mismatches:	231
Query Match:	5.04%	Indels:	140
DB:	1	Gaps:	20

US-09-802-285A-2 (1-659) x AB007344 (1-14385)

Qy	71	SerLysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnPro	90
Db	12733	TCAAAGACTATGAGAAGTCAAGAGTCCCTTGGAACGCTTGATGGACAATCGTTTTATG	12792
Qy	91	IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis	110
Db	12793	TTTGATAGT---CCTTGGGATATGGAGCTTGTTCAAATAATCCATAATCCAGCGC---	12846
Qy	111	LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys	130
Db	12847	-----ATGGTATGGGATCAAGTATTTGAAGT	12873
Qy	131	AspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuVal	150
Db	12874	GATCAGAAATGGTCTTATATGCTCAATCGACAAGAAATATCTTCGAGTTTATGATAGGG	12933
Qy	151	TyrHisAlaThrGlyAspGluLysTyrAlaArg-----GluTrp	163
Db	12934	TATCTGTGAGAAGGAGATAAGGACTATATTCAAAGTCAAGTCTCTTTCTATTGATGG	12993
Qy	164	ValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAsp	183
Db	12994	ATTGAGCAGGTGAGAGAATTTCTCCTCAATCCTTGATCACTAGAACCTTGATACGGGT	13053
Qy	184	-----LysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProPro	201
Db	13054	ATTCGTTCTTTACTTGGTGGAACTA-----	13080
Qy	202	ThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeu	221
Db	13081	---CTCTTGCTCCTCTTGAATTTGACTTGCTAGAGGAGAAAGAACTAGAGAAATTTTG	13137
Qy	222	AsnSerTyrHisGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln-----	238
Db	13138	GTCAGTCTAGAAAAGCAATGACTTTTATGAAGACTACTACGCGCCCAAGTACACCTT	13197
Qy	239	GlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGlu	258
Db	13198	AGTAATCGGGGATTTTACAAACAATTCGATGCTTGCTATCTATCTATCTTTTCAGAT	13257
Qy	259	PheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluLys	278
Db	13258	AAGATGGATCTAGAGAAGCTTACCATTTTGCTTCAGAGGAGTTGAAAACAGCAAAATTCAG	13317
Qy	279	LysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisVal-----	296
Db	13318	ACACAGATTTTAGGAGATGGAAGCCAGCTTTGAACAGTCAATCTCTATCATGTAGAGGTT	13377

Qy	297	-----AlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu	314
Db	13378	TATAAAGCCTTCGCTGGATTTTGTCCTC-----TTG	13407
Qy	315	GluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle	334
Db	13408	CTTCAGACTTCCAAGATAGTTACCAAGAGTTGCTGGAAAGATGGCGACCTATATTCAA	13467
Qy	335	SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLys	354
Db	13468	ATGATGACAGGCTTAGATGGACGACTTTGGCTTTTGGTGATAGCGATTCTACAGAAACG	13527
Qy	355	AsnPheArgMetAlaGlnPheAla-----	362
Db	13528	ACAGAAATTTTGAGCCTCTGCTGCTGTTTGAACAGGAGACCTCTTAAACGCTCTG	13587
Qy	363	-----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPhe	377
Db	13588	GATGTTAAAGTTGATTTGCTTAGCCTCTGTTCTGGGGCGAGAAAGGTCAAGCGACTG	13647
Qy	378	AlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAla	397
Db	13648	CAGGAATTTGAAAAGAGAGCTTGGCAGCCTAAG-----TCCATGATCTTTGAAGACTCT	13701
Qy	398	GlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAla	417
Db	13702	GGACATGCTCGATTAAAGAT-----GAACATCGTTATCTATTTTCAAAAAT	13749
Qy	418	SerProProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheLys	437
Db	13750	GGTCCGCTAGGAAGTGCCTATAGTCAGCAGAAATAGTTTTTGTCTTACAGTATCAA	13809
Qy	438	GlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMet	457
Db	13810	GGCCAACTATTTTCATAGATGCTGGCGGTTATTCTTATCG--GGAGATATATGAACGTTA	13868
Qy	458	LysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGln	477
Db	13869	TCTCTTAAAGAGTCTTGAGTCAATCCACCTGCATTGT-----	13907
Qy	478	AsnMetValIleThrLysAlaArgGlnAsnLysTrpGlu---ThrGlyAsnAsnLeuAsp	496
Db	13908	-----AGATGGAAAGCTCCGGAAGAAATCACCGG	13937
Qy	497	ValLeu-ThrTyr-----	500
Db	13938	ATCCTGGGAATATGAATACTATCCTCACTCCTCTGTTTGTCAACATAAAGAAAGGAGGG	13997
Qy	501	-----ThrAsnProSerTyrProAsnLeuAspHi	510
Db	13998	AGTGCATTATATTGAGGGGCTTATTGTCAGCAGAACTGATTTGCTTATCTTT---CA	14054
Qy	510	sGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAsp-----ArgAla	528
Db	14055	CAGGAGAAAAATCCTCATGTTGGTAGAGGATGTCGGCTCTGGTAGATGACATCAGGTG	14114
Qy	528	alleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnPr	548
Db	14115	TCAAGGTCAA-----CATGAGGTGTGACTCAGTTT-----	14145
Qy	548	oValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGly-----AsnAs	566
Db	14146	-ATCCTTCACAGGAT-----GTGACCTTCAAGATGGGAAATCAATCA	14189
Qy	566	nLeuMetIleGlnSer	571
Db	14190	GTTGAGACTATGGAGT	14205

RESULT 12

AF498414

LOCUS

DEFINITION Pseudomonas aeruginosa serotype O3 putative O-antigen biosynthesis

AF498414 19295 bp DNA linear BCT 13-JUN-2002

```

gene cluster, partial sequence.
AF498414 AC104733
AF498414.1 GI:20560007
Pseudomonas aeruginosa.
Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas
1 (bases 1 to 19295)
Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kutayavin,T.V.,
Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.
Genetic variation at the O-antigen biosynthetic locus in
Pseudomonas aeruginosa
J. Bacteriol. 184 (13), 3614-3622 (2002)
22053227
12057956
2 (bases 1 to 19295)
Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kutayavin,T.V.,
Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.
Direct Submission
Submitted (04-APR-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98105-2145, USA
On May 13, 2002 this sequence version replaced gi:17975262.
Location/Qualifiers
1..19295
/organism="Pseudomonas aeruginosa"
/serotype="03"
/db_xref="taxon:287"
<1..104
/note="ORF_1; C-terminal coding region of rpsA"
/codon_start=3
/transl_table=11
/product="RpsA"
/protein_id="AAM27759.1"
/db_xref="GI:20560008"
/translation="EKDAMELRKQEVESAGETTTIGDLIRAQWENQG"
241..525
/note="ORF_2; himd/ihfB"
/codon_start=1
/transl_table=11
/product="Himd"
/protein_id="AAM27760.1"
/db_xref="GI:20560009"
/translation="WTKSLIRIVTHQQLSAKQVLAIKTWLEQMSQALATGDRIE
IRGSGFSLHYRAPRGPNKPTGESVRLDGKVPHPKPKELDRVNEPE"
559..861
/note="ORF_3; hypothetical orfA"
/codon_start=1
/transl_table=11
/protein_id="AAM27761.1"
/db_xref="GI:20560010"
/translation="MQWFRSAVAVLIVSVAVVVLVLENGQVQLAFLGVQFPQWP
LAVYIIILGLGLAAQLPFLTISRVRASLRALVQARKEVDTLRAEPSKVS"
1146..2174
/note="ORF_4; wzz; similar to chain length determinant
protein"
/codon_start=1
/transl_table=11
/product="Wzz"
/protein_id="AAM27762.1"
/db_xref="GI:20560011"
/translation="WSKRNEIDFIELIQGLMAQKWLVLSTFLGVLAGSYAFLSKPV
YEAVAVLPPSLDVGAFNIAPSRDGLPFTVQSVYVDFRHDPEQAQWAKQYIELA
LPSLERSVSVGLRAFSDDSVSNQPTKELPERYTVDFRHDPEQAQWAKQYIELA
AQSDMDENQAQRELEVSRNIQQQIDTLRQVRRRTDRLVRKEALQVAEAVGVE
KPELIGQDQSLAISMEGLSMYMRGSRALAEIALEARTSDDPFIPALRGLEQLA
LYESKQLSPDRVAVFRQDGVETPDDPIKPRKVLILALGILAGGVLGLAALLRLER
EKRLDITG"
2421..3491
/note="ORF_5; similar to UDP-N-acetylglucosamine
2-epimerase"
/codon_start=1
/transl_table=11

```

```

/protein_id="AAM27763.1"
/db_xref="GI:20560012"
/translation="MAPVTALAGDGRFPEARVCVTGQHREMLDQVILNLFKIVDPDLN
IMRTGDLTDITVAILQGMKEVFAQKPDIIILVHGDTATTLAAASLAAYHQIPIVAHVE
AGLRTCMYSPWPEGRKLTGALACLTGAPTEKTLNRRREGIAAEQIIVTGNVTVID
ALLSVQRULHDVILARQAAPTEFLDSSRKFLVLTGHRSESGFERICRALAEVA
RNPEDIVVPHLNPPEPNRLSGVENIHLLEPLDLPFLHLSRAHILITDSG
GQCEAPSLGKVLVWRETERPEAVAAGTVKLIVTEVKNIVNELERLLTDESALQM
SFAHPYGDGEACQRIVQALLV"
3523..4785
/note="ORF_6; similar to UDP-glucose/GDP-mannose
dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="AAM27764.1"
/db_xref="GI:20560013"
/translation="MSFQTSIVIGLYIGLTAAVPAFRQKVKVIGVDVNOAAVDITNR
GHIHTVEPLDMVHAATVEGYLRATINPEPADAFLIAVETPKHEHEDLTFTIEAS
KSIAPVLKGDVILESTPGVATGMAELAEFRPDLTFPQTHGEASDIRVAHCFER
VLPGHVRELVDNRVIGMSKSEASRLKIFVKGECIITNATAKCMKLTENSF
RDVNTAFANELSIICKLDINWELIRLANRHPNVILQFGVGGHCHIAVDNPFVTS
KTPDEARLIRSAREVNDHKPFWLEKILAVADVLQKPEKTARDVSACFGIAPKAN
IDDLRESPAIAAQVVAHPGRVLVPEPNIELPGKFSGNLTLASTDSAMQEAADVW
LLVDHKEFREIAHPHRLQVLVDTRGLWEA"
5040..6155
/note="ORF_7; similar to Polysaccharide biosynthesis
protein; potential multiple membrane spanning domains."
/codon_start=1
/transl_table=11
/protein_id="AAM27765.1"
/db_xref="GI:20560014"
/translation="MLINALVGYASIFDAGLKSVVROIAISSAEKEKGPILGSSIL
FVLSGGLAALVHFAEYLVVDIFKVDKRSSTEDALTGRVAALCIPFELASLVLOGY
LEGVEDFYFNKRALSGTLTYLLPVAFLLFKTFSSLIYGLLFARITAGELVIAAP
RRCVMTWRMDVKFKTLISFGSLTITNIIISPMVMYMDKFLIARITAGELVIAAP
SELINRMSILPVAITRAVPPRLSALGQGSASVKRQALLYSGLVLPILVILFFVAPW
SELINRMSILPVAITRAVPPRLSALGQGSASVKRQALLYSGLVLPILVILFFVAPW
LMLYFSQQQGVVGAASISWLRMAVDCILMAYLDSK"
6162..7412
/note="ORF_8; potential multiple membrane spanning
domains."
/codon_start=1
/transl_table=11
/protein_id="AAM27766.1"
/db_xref="GI:20560015"
/translation="MASYIEKSLKFLFVLLTLMSPFKSAVLIKVILAFMFSKMF
LHVLGGVRRFSLDIIFFYLVILLAVILVGLNDGVAIALGDAFRILVVFSLIFL
VLYLWLENFDYLPILFSSIMVSTIIIFAPCLISFEALYSVSLFDFLFRVLNVAGRA
HEGHFQLGNNVSNLSLFTVPPLLSALVCSSELSKRYFLFVSLVGLVAVLLAGRA
LMLIILMVPFLAVFERRISGVSSIEGSGVGKLIIEVTFVAVCVGGVILLAGLVNQA
AERISARAGDSVRVDQFISLMWKFDTSPLGSGFGGSDVIRSEERPMWIELSTMQL
LFNGLLGSLITGICGFFYKALFVIRNFSLYRABAI STLGVLFVFCVGSATNPFYS
GDFSILFLGFIPLIASVGKVAHS"
7399..8463
/note="ORF_9; similar to Glycosyl transferase"
/codon_start=1
/transl_table=11
/protein_id="AAM27767.1"
/db_xref="GI:20560016"
/translation="WRIEASMLKSNRVDPIVLVLLNMWQDTIECLSLRHLQPOQ
DILVFDNGSENDVYQLELYRTISIAQHEEAFFELGDHVLHIVHDVSGIVVRLCSIG
ENLGSACNLAVKVAESMSVYVALLNNDTVLESOSIRIMHEVSDSDADIVICQIR
YDREPTIINWCGEVSRRWRVYFSGKNGVILEIKFFVPGFCCMLFKVDFERS
IGDFEIQFFGEEDVELSFRLEGRTHVVCARAIYHKVGASITGKDELIKRKTFIH
YLNRIINRITQLPVAVNWARFLVFKVAITLRKNFGKSSGIIAIFIRILLMQSSRLP
GVDKALFVLSKGLDYNDG"
8456..9469
/note="ORF_10; similar to Glycosyl transferases group 1"
/codon_start=1
/transl_table=11
/protein_id="AAM27768.1"
/db_xref="GI:20560017"
/translation="WVKAGRIIVIGGFFDPVGGVTAFSRLAENNMVSEVIDIYFSP

```

SKVPSPEKGRVVRKFTGFLSYLVANHROWAGRLHFNSTSRRAIVFFLVLKPFAS
 BESLMLHGSQVKSRLKMLYEALSLRDLYVFCNKEGRHQFVGRGLVPEEKWRLVSSY
 IKLPLSLIVBERKDIQEFDFGSRVLVAGSPSSLYNHEWCLIDYVSAREGKLALVSY
 GSGDALNRIBERANRDIKFIWNCQATFNHALSRASVILRPTSKDSFGVAVADAVEL
 GVPVVASVCKRYPGAGLPVDDFSAFCFTVDCILNGTDIIRTEDESGFRATFSYSLDL
 RDV"

CDS

9466..11607 /note="ORF 11; similar to Zinc-binding dehydrogenases"

/codon_start=1

/transl_table=11

/protein_id="AA027769.1"

/db_xref="GI:20560018"

/translation="MKQVLDMAKGGTLTVBAPAPQASKEHIVINSVLSLSAGTERM
 LVGPGKASVLEKARQOPEKVMLEKQVTDGLMTTIEAVRSKLGLOPLGVCNWCVUS
 EVGSGVEGKGVDRVVSNGPHADVVKSNLCARIPDEDDSAFVVFVVSIGLGIR
 LAQPTLGAFAVVTGVLGLLTVQLLRHAGKRVLAIDVDENKLSIARGVAGTGNPGR
 GEDPTAMATSRGGQVDGVIITASTQSDPVTAAHMRKRGRIVLVGVVTGLELNR
 DFYKELTFQVSCSPGPDYDENEKQDYPFGFVWTEQRNFPAILEDMLACGLD
 QALITHREAFEDAPAOYALVDDKEAGFHTLATAGGIVNGHOGKAEASTVDAN
 RPLVGFICAGNVASMLIPAKAQAQFHTLATAGGIVNGHOGKAEASTVDAN
 LANPTINTVAVVTRHSHARFVSQVLRAGKSVFVEKPLAIHAGELQEVRYDAQANSN
 GASALMVGPNVRFRFPQVKMSLLAAVKEPKSFMTWNAAGAIPIANHWTDKAGGGR
 IIGEACHFIDMLRAGSRIVSVQARSMDGAGVETEDKASITIFDFGDFGFTLLYL
 ANGAASFPERVEFTAGRVLQLDNFRFKGYGWPFGFKLNLWKQDKQNAKAKAFD
 GLQSQQAIPABEIEFVADVITIQVAELLRNQ"

CDS

11604..13448 /note="ORF 12"

/codon_start=1

/transl_table=11

/protein_id="AA027770.1"

/db_xref="GI:20560019"

/translation="MSSFMVKARTAWALGAPNLLVLAVMGVMTGLNPVRRLOBRAP
 SGDFPLTPPSPSVSLDKTPWLGKRVGYVYADSDVPMDSNLLTGVMNTSDPNW
 WEIPDFDLAVGDIKAVWEASRFDVPAALQKALGESALDKLNALWLDWCMNPPYK
 GPNWQGEASLRVWMLALALQVTPAPPALDLRLRHLRILRIVPTIRYALAQDN
 HGTSAALFTGGVSLRAGEPDVITWMLGRKWLLENRAKHLIGEDFSQVSYNIHR
 WMLDTFSLAEWRLQSLVFSARSARYSASATWHLHAFIDPATGDPNGLANDGARL
 LPLTQTDVRYPAVOLAMALFTDNTNAYGQBCSRPLLWGLDLPASBSVPVRSRVF
 DEGGVAILRSGQAMAMLVPRPRPSPQADALHLDLWLEGRNLADACTYSNTERSA
 LSYFPGTASHNTIQDGRDMPRLSRFLFGDWLAKATQAVPLTSSERASEMARAFSAAT
 DYGARHREVALTFDGLKVIDEISGSASVLRWLRPGSVWLDGNDLIGERGVLLI
 FESSMLLRILITIGWESRYLAKEEVPIEVIEVQAGTLKTDYEWAR"

CDS

13448..14662 /note="ORF 13; similar to Glycosyl transferases group 1"

/codon_start=1

/transl_table=11

/protein_id="AA027771.1"

/db_xref="GI:20560020"

/translation="MRVITYPHQHFSTPKGMVGIRSYEMARLVARGHQVWVWCGSYGG
 GOTGLDVPFSGRRQGTVDGIDIEFDLAYSNTDGFIKRTTTFPSFALRSIRLAWTEE
 YDLVATTPLTLAGIPGIFARWLKAFVFEVRLDLPKAMGVITNPVVLGMSVL
 EWLVSRAHRLIGLSPGIVGIERLGVPAHLVEMVTPNGCDLIDIPASDSEPRPERVKA

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0.000634	176.00	19295	137
Percent Similarity:	35.81%	Conservative:	80
Best Local Similarity:	22.61%	Mismatches:	244
Query Match:	5.04%	Indels:	147
DB:	1	Gaps:	33

US-09-802-285A-2 (1-659) x AF498414 (1-19295)

QY	113	TyrGlyTyrPheAspTyr-----	-----118
DB	11817	TTCCGCTACGTCCTTATGCCCTGATTCCTCGGTTCGGATGGCACAGTAAATCTGCTGACT	11876
QY	119	-----GlyLysAspIleAsnTrpGlnMetTrpProValIysAspAsnGlu	133
DB	11877	CGAGTGGGAATGACCAATCTCTGATCGAATTGG-----TGGGAGATCCCGATTTCGAT	11930
QY	134	ValArg-----TrpGlnLeuHisArgValIysTrpTrpGlnAla	146

DB	11931	CTGGCAGTGGGAGACATCAAGGCTGCTGGGAGGCGTCCGCTTCGACTGG-----GTG	11984
QY	147	MetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGlu-----	162
DB	11985	CCTGCTTTGGGCGCAAAAAGCTCTGGCTGGTGAGAGTGGTGGCTGGACAAACTGAATGCC	12044
QY	163	TrpValGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsn	182
DB	12045	TGGTTG-----AACGACTGGTGATGAATAATCTCTTACAAAGGGCCGGAAT-----	12092
QY	183	AspLysPheValTrpArg---ProLeuGluValSerAspArgValGlnSerLeuProPro	201
DB	12093	-----TGGAAATCGGCCAGGAGGCTTCGATCAGAGTCATGCACCTGGCCCTG	12140
QY	202	ThrPheSerLeuPhe-----ValAsnSerProAlaPheThrProAlaPheLeuMetGlu	219
DB	12141	GGGGCCATTCTGCTGACGACGAGTGCACCTCG-----CCTGCCGCACTGCTCGAC	12191
QY	220	PheLeuAsnSerTyrHisGlnAlaAspTyrLeuSerThrHisTyr-----AlaGlu	237
DB	12192	TTGCTGGCG--ATCATCTTCGAAGGATCGTGCACCAACCATTCGCTATCGGATAGCTCAG	12248
QY	238	GlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPhePro	257
DB	12249	GACAAACAATCACCGTACCTCCGAGGCGCAGCG---TTGTTTCATTGGTGGCAGTTGGCTT	12305
QY	258	GlupPheLysAspSerPro-----ArgTrpArgGlnThrGlyLeuSerValLeuAsn	274
DB	12306	CGGGCTCAGCGGAGCCCTGACGCCCTTACCTGGATGCGGTTAGGGAGAAAATGGCTGGAA	12365
QY	275	ThrGluIleLysLysGlnValTyrAlaAspGlyMetGlnPheLeuSerProIleTyr	294
DB	12366	AACCGGGCGAAACATCTGATTGGCGAGGAGCGTACTTCAGCCAGTATTCGCTGAATAC	12425
QY	295	HisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu	314
DB	12426	CACCGGGTCATGCTCGATACGTTCTCACTTGGCAGAGTGTGGAGGCGTTCAGTTTCCCTT	12485
QY	315	GluLysGluPheProGlnSerTyrValGlnThr-ValGluAsnMetIleMetAlaLeuI	334
DB	12486	GTGGAATTCCTGCTCGCTGGTATTTCGGGGCAAGTCAGCGACACTCCTGGCTGCACGCA	12545
QY	334	eSerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSer-----	349
DB	12546	TTTCATTCACCTGCTCCACT--GGTGACGGTCCGACCTTCGGGGTAAACGATGGTGCAGCT	12604
QY	350	-----TrpIleThrAspLysAsnPheArg-----Metal	359
DB	12605	GCTACCTCTGACCCAGACCGGATTCGTCGACTATCGACTCGCTGCGGTGCAGTTGGCCATGC	12664
QY	359	acGlnPheAlaSerTrpAlaArgValPheProAlaAsnGln-----AlaIleIle	375
DB	12665	ATTGTTTACCGAT---ACCAAGCCCTATCCCGGCAATCAAGAGTCAGCCGACCGCTGCT	12721
QY	375	sTyrPheAlaThrAspGlyLysGlnGlyAlaProAsnPheLeuSerLysLeuAlaLeuSe	395
DB	12722	GTGGCTGGGGCTCGATCTGCTTCGCCGAGTCTCTCTGTACCCGAGGCGGAGTGTTCGA	12781
QY	395	rAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetVal	415
DB	12782	CGAGGGTGGCTACGCCATTCCTGCGAGCGGG-----CAGGCAATCGCAAT	12826
QY	415	ulysAlaSerProProGlyGluPheHis---AlaGlnProAspAsnGlyThrPheGluLe	434
DB	12827	GCTGGCTTACCCAGGTTCCGTTTCGCCCCAGCAGCGGAGTGTCTTCATCTTGAATCT	12886
QY	434	upheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspG	454
DB	12887	ITGGCTCGAAGGGCGAAATCTGCTCGCGGATCGCCGCACTACAGCTACACACCGAATC	12946
QY	454	uAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLe	474
DB	12947	CGCTTGGCTG-----AGTTACTTCCCGGAACGGCCAGTCACACACGATCAGTT	12997

QY	474	uAspAsnGlnAsnMetVal-----IleThrLysAlaArgGlnAsnLysTrpGluThrG1	492
QY	474	uAspAsnGlnAsnMetVal-----IleThrLysAlaArgGlnAsnLysTrpGluThrG1	492
Db	12998	CGATGTCGCGATPCAGATCCCGCTCAGCCGCTTCTTTTGGCGACTGGTGAAGGC	13057
QY	492	yAsnAsnLeuAspValLeuThr-----	499
Db	13058	GAGCAGGCTGTCGCCCTGACTTCGGAGCGAGCGTCGGAGGAATGAATGATTTTCGGC	13117
QY	500	-----TyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerVal-----LeuPheI1	517
Db	13118	GGCCTATACCGAT---GGGTAT---GGGCCAGGCATTAAGCGGGAGGTTCCTTGTTC	13171
QY	517	eAsnLysLysTyrPheLeuValIleAspArgAlaIleGlyGluAlaThr-----	533
Db	13172	GGATGGA-----CTCAAGGTTATCGATGAGATTTCCGCGCTTTCCGCGAGTGCCTGCT	13225
QY	534	-----GlyAsnLeuGlyValHisTrp-----GlnLeuLysG1	544
Db	13226	GCCTGAGCGCTGGAGCCTGGAG-CTGGGTGTCGATGGAACACGCTCAAGGAGAGA	13284
QY	544	uAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrTyrArgAspG1	564
Db	13285	GGGAGACGCTGCTGATTTTCGAAAGCTCAT-----GCCTCTTCGGTATCGAGCTGA	13338
QY	564	yAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArg-----	576
Db	13339	CAATCGGGTGGAGTCGCGCTACTACTTCGAAAGAGGAGGTTCTCTCATCGAGATAG	13398
QY	577	-----ThrSerLeuAsnGluGluGluGly-----LysValSerTy	588
Db	13399	AGGTTGGGAGCGCGGAAACATTCAGACGGATACAGGTGGGCCCGATGAGAGTACTTA	13458
QY	588	rValTyrAsnLysGluLeuLysArgProAlaPheValPheGluLysProLysLysAsnAl	608
Db	13459	TTTTTCATCAACAC-----TTTTTCGATCCGAAGGAGGATCGT	13494
QY	608	aglyThrGlnAsnPhe	613
Db	13495	AGGCATCCGCTCCTAC	13510
RESULT	13		
AF498401			
LOCUS	AF498401	16720 bp	DNA linear BCT 13-JUN-2002
DEFINITION	Pseudomonas aeruginosa serotype 010 putative O-antigen biosynthesis		
ACCESSION	AF498401	AC104720	gene cluster, partial sequence.
VERSION	AF498401.1	GI:20559773	
KEYWORDS			
SOURCE	Pseudomonas aeruginosa.		
ORGANISM	Pseudomonas aeruginosa		
REFERENCE	1	(bases 1 to 16720)	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
AUTHORS	Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kutayavin,T.V.,		
TITLE	Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.		
JOURNAL	Genetic variation at the O-antigen biosynthetic locus in		
MEDLINE	Pseudomonas aeruginosa		
PUBMED	J. Bacteriol. 184 (13), 3614-3622 (2002)		
REFERENCE	2	(bases 1 to 16720)	
AUTHORS	Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kutayavin,T.V.,		
TITLE	Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.		
JOURNAL	Direct Submission		
COMMENT	Submitted (04-APR-2002) Genome Center, University of Washington,		
FEATURES	Box 352145, Seattle, WA 98105-2145, USA		
source	On May 13, 2002 this sequence version replaced gi:117975249.		
	Location/Qualifiers		
	1. .16720		
	/organism="Pseudomonas aeruginosa"		
	/serotype="010"		
	/db_xref="taxon:287"		

/transl_table=11
/protein_id="AA027559.1"
/db_xref="GI:20559780"
/translation="MLVKLNENSVNLSVAIQANALLPLLAFFYLPGLVLEQCAFAR
LVAAALAVYLVITCLYSFDISGVQVIVDAKEEREVEAKFFNFPAARVGLLIIFSSVP
LVGGYLFPMGADLVVITVLFVLMILQSNYYFAVENWNLATFVSSRTAVLAV
YFNGKSTVTLFAGSLVGGAVGVLTRFGLSEFKRIRGDDVVVLLWSGRH
LVFLGSLVTLFRGANVILLAGVNSAAVAVASEKIISTQALARPUNOLFPLKVVK
AWASLERERNAPAFRLVKNKTAVOILLMVFLPVLFIIFYFCWVLGVIPGESEKVF
FLVIMSPAVIPGVANSMFGAVGLSLGGQYFYFAYSIIIGVCAFLSLVMSYLLSDT
GAAMTVLAEALLFAFLCKYGRPRHG"

CDS

6170..7504
/note="ORF_8; similar to u-PAR/ly-6 domain; potential
multiple membrane spanning domains."

/codon_start=1
/transl_table=11
/protein_id="AA027560.1"
/db_xref="GI:20559781"

/translation="MDEFYILKINIFLSSFFLLGAMCARKIFSVWLNPNVIFISLLWF
FYTVPLLVAFEPVSPFAILYILIFCFVSLGSFVFPWVAFGRNENKPNVASCYFV
PLMVAFYSCSISAIMIFWGLQGLSLEOLLNPNVSGIGYIAGKRYSGRIVSVFS
OVGLQCSYTAFFGLIYGARSEARGPKLLFFTLPALVWLOKSAKGLFFPSIFLIF
FGMLGRVNNKNTILVGFDFRLLYGLLILPLIFPLSRGLYQLDDTALMINRL
RYXIVTSSVHLPAFSDWFSERYGESLMHYRQSSSTLGFYFMSFPLQAGDSDVAM
GVYDFEFTYGEYVNGNYTVPRGIITDFGLLSGLFFAFILGLFNCVRLCRNGSA
FAILFCYFVAISYQTYIISLTLWTLTPFVIVQWFMGLFLKFRFYTRSKV"

CDS

7501..8343
/note="ORF_9; similar to Glycosyl transferase"

/codon_start=1
/transl_table=11
/protein_id="AA027561.1"
/db_xref="GI:20559782"

/translation="MKICLPVCVNAHDDAVRELDVSNTAFSCGLGVLELYLSDNST
LSSVITDARNYSNRYLQNSNIGYPPAPARALSTLTPEIDSDVFMVNCVNDLVSK
DFFIALSSVCKDVIAPGIFSKDGRDNPIMRPSLRKMKFMVWCSSTALFR
WYMLVARERTRACTORKIAKGGAIGQPAMYGAGSGFIITQLYPKRGAHLYPR
FLFGEYFVAEQLNNMLIQHPDVRVSDKEHASTSKIKLFCPEHRKSYEYLLSK
FFDD"

CDS

8457..10625
/note="ORF_10; similar to Zinc-binding dehydrogenases"

/codon_start=1
/transl_table=11
/protein_id="AA027562.1"
/db_xref="GI:20559783"

/translation="MFCIEGAMQVLQDMAKGTSTVAPQISRHDVILNVTSL
ISAGTERMLVGFGLYDLKARQOEKVMVLEKIKTDGLMTVEAVQSKLAQPLGL
YCNVGVYSEIGSGVEGPKVGRVNSGPHADLVKVKNLCAIRPDGVDDESAFVVA
SIGLQIRLAQPTLGEFCFVTVGVLIGLITVQLRAHGRVLIADPDDSKULELARQY
AQTCNPKGEDVPAAGMAFSKGVGVIIITASTKSSDPVTQAAARMKRGRIVLVGV
VGLNRADFEYKELTQVSGYGRDYDEYBKGGDPLGVFRWTEQRNFEAVLDM
LDSQLEVKTLITHRFKPEDAPAAYSTLTDONSGLMLLYASDTAERVVRVPLSDM
KGFNAQPVVGFICAGNYASRILIPAFKTAGAOPHTLTSGGINSVILHGEKAGFSQAS
TNIDAVNERMINITAIIVTRHDTHARLVVDALRTGKNVFEKPLIAITAELEDVQAH
AITHQTGGPQLMGVFNRRFAPHIQMKGLLASVREKPSFINTNAGAPVGVHWTQDI
AVGGRIIGEACHPIDLMRYLAGSRIVSVQAPRMGDSNAIDTSDKASITLGPEDGSF
GTILYLANGASFPKRIEYFTAGRVQLDNRKILKGYWPGFPGFNKMLWKQDKQRQC
VAFLDSTIEKGEPTIACEEIIFEVTRVSEIAEAELRNQWD"

CDS

10634..12286
/note="ORF_11; similar to Prenyltransferase and squalene
oxidase"

/codon_start=1
/transl_table=11
/protein_id="AA027563.1"
/db_xref="GI:20559784"

/translation="MALKVRLVHTLYRLKQIIFEVYVFAQVRVHYLYRVAVR
WERDSAPSRLNSNDAYVEFLGIGTKVQCQPDWQADRSKLLWNLHYLDDLNAR
DIGQPGLAQDLQISWIQANPPVSGEGVEPYLSIRIVNLVKLARHDERSTFLADSL
AVQADALQVEXHILGNHLSFANGKALVFAGYLSGAMADRWLAKGLILDEELPEQF
LNDGGRFELSPMYIATLWMDNLSLRSGLEDLAERLPQWREVVVQGLKRLSMQ
HPDGRIISFNDAAFIAPEDYIAAYAKRLDISPPAHENHLAALYNAGVAAVLPA
GKAILDLAKGPDYOPGHAHADTLPSELSVFGKRLVNSGTSOYGDSDSRQORGTA
AHNTVGLLGYSDSSEWAGFVRARAAVIERFEVEPRITCTQASHDGVGLYKIRHR
RTWTFTRTSMEIFDSITGDQVATSRFPFSEVKVMKADGFVADMGKVKVSIIDFSG

SDDVRLISSTWHPFGCGLANQCLVATFSGDSLTRICWGGI"
12355..13467
/note="ORF_12; similar to Glycosyl transferases group 1"
/codon_start=1
/transl_table=11
/protein_id="AA027564.1"
/db_xref="GI:20559785"
/translation="MSALGALPENSOLEVITTPNRYSSPLEAPLLEKHSRLIIHR
IPFOHNSGMLDOSRAFPYSLKVLGTRNKDQDVVATSRSLTASIGALVABQLRC
PLYLIDRIDFVDITIKDLPKRSIFVFPFISMERIIVRTATRVNLVAGFLPFNSR
YPGDFFVFTINGIDDEFKVQTSCEBESSSLYTLVYAGNMGEGGHLIILPQIARHF
DHLRFLVSGIKFGERKQDAIEAGCTNVEITSPVPSSELIAAYQADVLFJLHNDY
DAFKFLPSKIFEYAAKQPIWAGVAGYAGFIQDFISNAAVFPCNVEDAIFKAFSSL
SMVTKPRFVFERFAREKIMKALAEIDIFVVRGEVK"
13464..14417
/note="ORF_13; similar to NAD dependent
epimerase/dehydratase family"
/codon_start=1
/transl_table=11
/protein_id="AA027565.1"
/db_xref="GI:20559786"
/translation="MRVFLGITGTVGLATSKLLQEGHCVRAAVRDANPALAQEVEQ
VFGNLVKHWDWASSLEQGVVHCAARVHMKDASRDLSEFRVNVEGTINLARQAA
KAGVKRFVFISSIKVNGESTRLGVPTADDIAMPFLDPYGVSKLEAEQKLLSARVDM

Alignment Scores:

Pred. No.:	0.0015	Length:	16720
Score:	170.50	Matches:	88
Percent Similarity:	36.61%	Conservative:	46
Best Local Similarity:	24.04%	Mismatches:	151
Query Match:	4.88%	Indels:	81
DB:	1	Gaps:	18

US-09-802-285A-2 (1-659) x AP498401 (1-16720)

Qy	144	TrpGlnAlaMetAlaLeuValTyHiseAlaThrGlyAspGluLysTyrAlaArgGluTrp	163
Db	10871	TGGCAAGCT-----GATAGGACGAGTAATAATG	10900
Qy	164	ValTyr-----GlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAsp	181
Db	10901	CTCTATAATCTGCACATATTTGGATGATCTGAAACGCTAGGACATTGGC---AGTCAGCCT	10957
Qy	182	Asn-----AspLysPheVal-----	186
Db	10958	GGGTAGCCGATAACTAAATCAAGTTGGATCCAGGCTAATCCACCGGTATCCGGTCAA	11017
Qy	187	---TrpArgProLeuGluValSerAspArgValGlnSerLeuProThrPheSerLeu	205
Db	11018	GGTTGGGAGCCCTATCTCTTCCTTCCTTCGATATCGTTAATCTTGTGAAATGGCTTGC	11077
Qy	206	PheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrrhis	225
Db	11078	CATGATGAGCGTTTCACCTTT-----CTGGCGGACAGTCTAGCT	11116
Qy	226	GlnGlnAlaAspTyrrLeu-----SerThrHisTyrrAlaGluGlnGlyAsnHisArgLeu	243
Db	11117	GTACAAGCGGACGCGCTGGTGCAACAAGTAGAATATCATATCTTGGGAAATCACTATTC	11176
Qy	244	PheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAspSerPro	263
Db	11177	GCTAATGGGAAGCGTGTGTGTTTGGCGGC---GCTTATCTGTCCAGCGCGATGCTGAT	11233
Qy	264	ArgTrpArgGlnThrGlyIleSerValIleAsnThrGluIleLysGlnValTyrrAla	283
Db	11234	CGTTGGCTTGCACAGGCTTACGAGTACTTGCAGAGAGCTCTCTGAGCAATTTCTGAAT	11293
Qy	284	AspGlyMetGlnPheGluLeuSerProIleTyrrHisValAlaIleAspIlePheLeu	303
Db	11294	GATGGTGGGCAATTTTGAATTTGCTCCGATGTATCACGCCACTTTTACTC-----	11341
Qy	304	LysAlaTyrrGlySerAlaLysArgValAsnLeuGlu-----	315

```
Db 11342 -----TGGATATGTCGACCTGGTGAATCTCTCCAGCGGTCCGGCTACCTGATTG 11395
QY 316 ---LysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle 334
Db 11396 GCTGAGCGTCTACCGCAA---TGGCAGAGGTTGTGTCACAGGATGAAGTGGCTTCGT 11452
QY 335 SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLys 354
Db 11453 AGCATGCAGCATCGGATCGTAGAATTAGCTTCTTCAATGAT----- 11494
QY 355 AsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIle 374
Db 11495 -----CCTGCATTCCGA-----ATTGCACCTGAATATGAAGATATT 11530
QY 375 LysTyrPheAla-----ThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLys 392
Db 11531 GCAGCGTACCCCAAGCGCTCGATATATCACCTCCAGCTCATGAACACCATCGCGCGG 11590
QY 393 AlaLeuSerAsnAla---GlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThr 411
Db 11591 ATTTATAATTCGGCTACAGTTATGCTGCTGTTACTTCTCCCGCATCGCGTGAAGTATA 11650
QY 412 ValMetValLeuLysAlaSerPro-----ProGlyGluPheHisAlaGlnProAsp 428
Db 11651 TTGGATCTTGCAGAGTCGCTCCGACTATCAGCCGGGA-----CATGCCATGCCGAC 11704
QY 429 AsnGlyThrPheGluLeuPheIleLysGlyArgAspPheThrProAspAlaGlyValPhe 448
Db 11705 ACATTGAGTTTGGATTAGCGCTTTGGAAAACGCTTAGTAGTCAATTCCCGAATTCT 11764
QY 449 ValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIle 468
Db 11765 CAGTACGGTGATGACTCTGAG-----CGAAGGAGCGGTGGTACTGCGCGG 11812
QY 469 HisSerThrLeuThrLeu 474
Db 11813 CATAATACGGTGGGATTG 11830

RESULT 14
AF498411
LOCUS AF498411
DEFINITION Pseudomonas aeruginosa serotype 019 putative O-antigen biosynthesis
ACCESSION AF498411
VERSION AC104730
KEYWORDS AF498411.1 GI:20559940
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa.
REFERENCE 1 (bases 1 to 16720)
AUTHORS Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kuttyavin,T.V.,
Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.
TITLE Genetic variation at the O-antigen biosynthetic locus in
Pseudomonas aeruginosa
JOURNAL J. Bacteriol. 184 (13), 3614-3622 (2002)
MEDLINE 22053227
PUBMED 12057956
REFERENCE 2 (bases 1 to 16720)
AUTHORS Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kuttyavin,T.V.,
Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98105-2145, USA
COMMENT On May 13, 2002 this sequence version replaced gi:17975259.
FEATURES
source
1..16720
/organism="Pseudomonas aeruginosa"
/serotype="019"
/db_xref="taxon:287"
<1..104
/note="ORF 1; C-terminal coding region of rpsA"
/codon_start=3
CDS
```

```
/transl_table=11
/product="RpsA"
/protein_id="AA027704.1"
/db_xref="GI:20559941"
/translation="EKDAMKELRKQEVESAGPTTIGDLIRAQWENQG"
241..525
/note="ORF 2; himD/ihfB"
/codon_start=1
/transl_table=11
/product="HimD"
/protein_id="AA027705.1"
/db_xref="GI:20559942"
/translation="WTKSELIRIVTHOGLSAKOVLEIAKTMLEQMSQALATGDRIE
IRGGSFSPSHYRAPVRGNPKTGESVRLDGKFPVHPKPKGKELDRVNRFE"
601..858
/note="ORF 3; hypothetical orfA"
/codon_start=1
/transl_table=11
/protein_id="AA027706.1"
/db_xref="GI:20559943"
/translation="WVSIWVVVFLVLENOGLVQVAFGLQSPQWPLAVYLITAFVLGGL
LGLAIQLPSLAISRARASGLRAELEQARKEVDSLVRPSSS"
1114..2196
/note="ORF 4; wzz; similar to chain length determinant
protein"
/codon_start=1
/transl_table=11
/product="Wzz"
/protein_id="AA027707.1"
/db_xref="GI:20559944"
/translation="WTDGIDSPQTSKRNEIDFIELLOGLWSOKWLVSLTFLGAVLAG
SVAFSLKPVYERVALPSPSLDVAGFNARGDSGLPFTFYQVTSVLRNQQSEGT
RKQFPREVLPSELEKSGIDGLYRAFSDVSVNRPKSLDPERYTVIVNRQDPEQAAK
WAKQYADLAARLSMDENRGAQRELEKSRNLIQDITLRQVVKERRNDRIVRLKEAL
QVARAVGEKPLIGAIQDOOLSAIMESGLMVRGTAKALEIKALKRKTSDPPTPA
LRGLQOLALYESKOLNDRVAVFRQDGGVEIPDDPIKPKKALLALGALAGGVFVGL
LALLRLLDGRNGMADLKHPMTDK"
2403..3488
/note="ORF 5; similar to UDP-N-acetylglucosamine
2-epimerase"
/codon_start=1
/transl_table=11
/protein_id="AA027708.1"
/db_xref="GI:20559945"
/translation="WAPVTALADDDREARVVCVTGQHREMLDQVNLNFKIIVDFDLN
INRSGDLTDVTVAIQGKQVFSQFKPDIVLVHGDITATLAASLAAYHQIPVAHVE
AGLRGTNMYSPWPERGKRLTGALACLFAPETGRDNLREGISAERIRVTGNTVID
ALLNVQRLQODMELASAAAAPTEFLDSRRFILVTHRRSEFGGGERICQALGVVA
RNHPEDIVYPVHLNPNVREPUNRLISDVQNIHLIEPLDLPFVHLSAHVILDTSG
GLOEAPSLGKPLVMRETERPEAVAAGTVKLVGTEIKNIVNELERLLVDSAYRKM
SFAHPYGGGWASKRIIDSLDAKRG"
3569..4837
/note="ORF 6; similar to UDP-glucose/GDP-mannose
dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="AA027709.1"
/db_xref="GI:20559946"
/translation="MSFTISVIGLYIGLPTAAVFASRRKQVIGVDVNSKAVETINQ
KAIHTEPDLMDLVHAATNGYLRVAVSPEDAFLIAVPTPFKEQHPDLRIYESAC
KALAPVLKAGDLVILESTSPVGCATQMSAQLSAERPDLTFFQSHGDSADIRAHCPER
VLPHVRLQVNDRVIGGMTATCSAAASLYKIPVEGECVTITNATAMCKLNSPF
ADTNIAFANELISICLDLINWELIRIANRHPVNLQIPQGVGGHGIADVDPWIFS
KTPGEARLIRAREVNDKGLVNLNKIEMVAEFLQANPRTAKDVTIACGLAFKPD
IDDLRESPAVSIAQNVKMSHPGTVLWPEPTAELPEKLAERLDLVSINSALERSEVIV
LLVDHKEFKFSPCDLGAYKVIDTRGVNA"
4915..6177
/note="ORF 7; similar to Polysaccharide biosynthesis
protein; potential multiple membrane spanning domains."
/codon_start=1
/transl_table=11
/protein_id="AA027710.1"
/db_xref="GI:20559947"
```


/translation="MLVKLNENSVNLSVALIOGANALLPLLAPPLFGLVLEQGAFAF
LVAALAFYVLTCLYXSFDSGVQIVDAKEEREVEAKFPFNFAARVGLLIPSSVF
LVGGTLFWGADVLVLLVFLVGLQILSNFYFOAVENWLLATFVSSRVAVIAV
YVGHKGDTDLASILVGLSGLVGGVGVGLTRFGLSEBRKRGDGVVLLWSGRH
LFLNLSVTLFRGNALVLLAGVNSAAVVALSEKI IKSTQALARPINQLFPRVKV
AWASLERNNAPFLRVNLTQAVLLMVFLLPFVFLFGLVGLVGLVPGFSEKVF
FLVIMSPAVLPGVANSMEGAVGLSLGQTVYFASVSIIGVCAFLISLVMSYLLSDT
GAAMTFVLAELLLFAFLCKYRGKPRHG"

CDS

6170. .7504
/note="ORF_8; similar to u-PAR/ly-6 domain; potential
multiple membrane spanning domains."

/codon_start=1
/transl_table=11
/protein_id="AAM27711.1"
/db_xref="GI:20559948"
/translation="MDEFFYLKLNIFLSPFLGAWCARKIFSWLNPVSIFFSLWP
FYTPFLVAPEAPVPAIILYIFCFVLSLSPFPGWAFSPNKNKSNACSYFDV
PLMWLAVSCGISAIMIFWGLQOQISLEQLNPVSIIGIYAGKRYSGEIVSVFVS
PVLGQCSYTVAVFGLIYGARSEARGLKLPFTLPALLVMLQSAKGLFFFSIFLF
QVGLVSRVKNVNTLVGKDFRLLYGLLILPLIFSLSRGIYQDDTDALMINRL
RYLVTYSVHLPAESDWPSEYFCESLMHYROESSTIGFTFMSFFQLAGDSDVAM
GYIDFFYGFVKNLYTVRNGIITDFGLGSLFFAFILGLFCMLCWRLLRNGSA
FAIFFCFYFAISYQTYIISLTLTLTPFVFIQVFMFLGLLKFFRYTRSKV"
7501. .8343
/note="ORF_9; similar to Glycosyl transferase"
/codon_start=1
/transl_table=11
/protein_id="AAM27712.1"
/db_xref="GI:20559949"
/translation="MKICLPVCYNADDDAVRFLDSVNTAFSCFGLVLEVLSDNST
LSSVTFDARNYSYNYRFLQNSNIGYFAPAFALSTLTPEDSIDYFMVNVLDLVSK
DFFIALLSPCKGVGLIAGPISDKDGRDLNPKIMRPSLRKMKPMVMVCSSTALFR
WYSLVABREBTRACTQRIAKKGAIQGPAMYGAHGSFIITOLYKRGHVLVPR
FLFGEVFAEQRLNNMLIQVDPVRSDKEHASTSKIKLKFICPEHRKSYEYLLSK
PFDD"

CDS

8457. .10625
/note="ORF_10; similar to Zinc-binding dehydrogenases"
/codon_start=1
/transl_table=11
/protein_id="AAM27713.1"
/db_xref="GI:20559950"
/translation="MSFCIEGAMQVLQDMARGGTSVTVAPAPQISRDIHINSTVSL
ISAGTERMLVGGKASYLDKARQPEKVMVLEKTKDGLMTTVAVQSKLAQPLPLG
YCNVGVSEIIGSGVGFVGVDRVNSPHADLVKVSKNLCARIPDVDDSEAAFPVVA
SIGLQILAPLQTLGECFVTVGVLITVQLLRAHGORVLAIDFDSKLELRQYG
AQTCPNGKEDFVAAGMFAFSKGVGVIIITASTKSDPVTQAARSKRKRIVLVGV
VGLNLEDFEYKELTFQVCSYGPGRVDPEYEKGDDYPLGFWKMTBOENPEVLDM
LDGQLEVKTLITHRFKPEDAPAYSTLTQDNSGLMGLLYASDSTAERYVRHVPLSDM
KGFNAQKPVGFIGAGNYASRLIPAFKTAGAQFHTLVTSNGINSVHGEKAGFSQAS
TNIDAVNERMINTAI VTRHDTARLVVDALRTGNVFEVEXPLAITHAELEDVQAVH
AITHOTGGSPOLMVGNRRFAPHIOKMGLLASVREPKSFIMTMNAGALPUGHWTQDI
AVGGRIIGEACHFIDMLRYLAGSRIVSVQARRMGDSNAIDTSEDKASITLGFEBGSP
GTILYLANGASSFPFERIEVFTAGRVLDNFRKLKGYGWPFGNKNLWKQDKGRQC
VAAFLDSIEKGEPTIACEBIEFVTRVSIIEAAEILRNQWD"
10634. .12286
/note="ORF_11; similar to Prenyltransferase and squalene
oxidase"
/codon_start=1
/transl_table=11
/protein_id="AAM27714.1"
/db_xref="GI:20559951"
/translation="NALQKLVRLYHTRLYRLKQIFRVVYFPAQVRVHYRNVAVRR
WRDWSAFSRNLSNDVAVEFTFLGITGVQVCPEDWADRSSKMLYNLHYLDDLNAR
DIGSQGLADLQSIQANPPVSGEGWEPYPLSRLIUNLVKARLHBERSTFLADSL
AVQADALVQOVBHYILGNHLFANGRLVFAAGLAYLSGAMADRWLAKRLRDLDELPEQF
LNDGGFELSPMHTLMLDMCDLVNLSRGLPOLAERLPQWRVGVVQGLKWSMQ
HPDGRISFNDAAFGAPEYDIAAYAKELDISPPAHENHLAAIYNSATGYAAVLPA
GVKALLDLAKVGPDPQPHAHADTLSELSVFGKRLVNSGTSQYCDSDSERQORQTA
TRVTLGLLGDYSSEVWAGVRRAAVVIERVEPEITICIOASHDGVRLVKGLRHR
RWTFTFRSMELFDSITGQVVATSRFFPESPEVKVMKADGFVADMGAGKKYSIDPSG
SDVRLIASSWHPFGCSLANQCLVATFSGDSLTITRICHGGI"

CDS

12355. .13467
/note="ORF_12; similar to Glycosyl transferases group 1"
/translation="MLVKLNENSVNLSVALIOGANALLPLLAPPLFGLVLEQGAFAF
LVAALAFYVLTCLYXSFDSGVQIVDAKEEREVEAKFPFNFAARVGLLIPSSVF
LVGGTLFWGADVLVLLVFLVGLQILSNFYFOAVENWLLATFVSSRVAVIAV
YVGHKGDTDLASILVGLSGLVGGVGVGLTRFGLSEBRKRGDGVVLLWSGRH
LFLNLSVTLFRGNALVLLAGVNSAAVVALSEKI IKSTQALARPINQLFPRVKV
AWASLERNNAPFLRVNLTQAVLLMVFLLPFVFLFGLVGLVGLVPGFSEKVF
FLVIMSPAVLPGVANSMEGAVGLSLGQTVYFASVSIIGVCAFLISLVMSYLLSDT
GAAMTFVLAELLLFAFLCKYRGKPRHG"

CDS

13467. .14417
/note="ORF_13; similar to NAD dependent
epimerase/dehydratase family"
/codon_start=1
/transl_table=11
/protein_id="AAM27715.1"
/db_xref="GI:20559952"
/translation="MSALLGALPENSQLEVIPTQPNRYSSFSLEAPLEKHSRLIHR
IPITQHNSGMLDQSRAFSGLYGLVITRNKDYDVVTATSSRLMTASGALVQRLRC
PLYLDIRIDFVDTIKDVLPSRFPKPIFSMERITVTRATRVNLVAGSLFYPNSR
YPGDDFVVTNGIDIDEFVKVQTSCEERESSLYTLVYAGNMGGGLHLIIPQIARHF
DGHRFLCIGDGRKQLQDAIBRAGCTNVTSPRSELIAAYQKADVLFLHNDY
DAFRPLPSKIFIEYALGKPLWAGVAGSAGFIQDITSNAAVFRPCNVEDALKAFSSL
SMVTKPRKFEVERFAREKIMKALADIIIFVVRGEV"
13464. .14417
/note="ORF_13; similar to NAD dependent
epimerase/dehydratase family"
/codon_start=1
/transl_table=11
/protein_id="AAM27716.1"
/db_xref="GI:20559953"
/translation="MRVPLTIGITGVGLATSKRLIQEGHCYRAAVRDANPALAORVEQ
VVFGLNVKHDWASSLEGOEVVHCAARVHVMDASRDLSEFRVNVVGTNLARQAA
KAGVKRFVFISSIKVNGESTRLGVPTADDIAMPDLPYGVSKLEAEQKLLSLAARDVM

CDS

14417. .16720
/note="ORF_14; similar to NAD dependent
epimerase/dehydratase family"
/codon_start=1
/transl_table=11
/protein_id="AAM27717.1"
/db_xref="GI:20559954"
/translation="MRVPLTIGITGVGLATSKRLIQEGHCYRAAVRDANPALAORVEQ
VVFGLNVKHDWASSLEGOEVVHCAARVHVMDASRDLSEFRVNVVGTNLARQAA
KAGVKRFVFISSIKVNGESTRLGVPTADDIAMPDLPYGVSKLEAEQKLLSLAARDVM

/codon_start=1
/transl_table=11
/protein_id="AAM27715.1"
/db_xref="GI:20559952"
/translation="MSALLGALPENSQLEVIPTQPNRYSSFSLEAPLEKHSRLIHR
IPITQHNSGMLDQSRAFSGLYGLVITRNKDYDVVTATSSRLMTASGALVQRLRC
PLYLDIRIDFVDTIKDVLPSRFPKPIFSMERITVTRATRVNLVAGSLFYPNSR
YPGDDFVVTNGIDIDEFVKVQTSCEERESSLYTLVYAGNMGGGLHLIIPQIARHF
DGHRFLCIGDGRKQLQDAIBRAGCTNVTSPRSELIAAYQKADVLFLHNDY
DAFRPLPSKIFIEYALGKPLWAGVAGSAGFIQDITSNAAVFRPCNVEDALKAFSSL
SMVTKPRKFEVERFAREKIMKALADIIIFVVRGEV"
13464. .14417
/note="ORF_13; similar to NAD dependent
epimerase/dehydratase family"
/codon_start=1
/transl_table=11
/protein_id="AAM27716.1"
/db_xref="GI:20559953"
/translation="MRVPLTIGITGVGLATSKRLIQEGHCYRAAVRDANPALAORVEQ
VVFGLNVKHDWASSLEGOEVVHCAARVHVMDASRDLSEFRVNVVGTNLARQAA
KAGVKRFVFISSIKVNGESTRLGVPTADDIAMPDLPYGVSKLEAEQKLLSLAARDVM

CDS

13464. .14417
/note="ORF_13; similar to NAD dependent
epimerase/dehydratase family"
/codon_start=1
/transl_table=11
/protein_id="AAM27716.1"
/db_xref="GI:20559953"
/translation="MRVPLTIGITGVGLATSKRLIQEGHCYRAAVRDANPALAORVEQ
VVFGLNVKHDWASSLEGOEVVHCAARVHVMDASRDLSEFRVNVVGTNLARQAA
KAGVKRFVFISSIKVNGESTRLGVPTADDIAMPDLPYGVSKLEAEQKLLSLAARDVM

Alignment Scores:

	Pred. No.:	Score:	Length:	Matches:
		0.0015	170.50	88
		Percent Similarity:	36.61%	Conservative: 46
		Best Local Similarity:	24.04%	Mismatches: 151
		Query Match:	4.88%	Indels: 81
		DB:	1	Gaps: 18

US-09-802-285a-2 (1-659) x AF498411 (1-16720)

QY	144	TrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTyrP	163
DB	10871	TGGCAAGCT-----GATGAGCAGTAAATATG	10900
QY	164	ValTyr-----GlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAsp	181
DB	10901	CTCTAATACTGCACCTATTGGATGATCTGAACGCTAGGACATTGGC---AGTCAGCCT	10957
QY	182	Asn-----AspLysPheVal-----	186
DB	10958	GGGCTAGCGGATAAACTAATTCAAAGTTGGATCCAGGCTAATCCACCGGTATCCGGTGAA	11017
QY	187	---TrpArgProLeuGluValSerAspArgValGlnSerLeuProThrPheSerLeu	205
DB	11018	GTTTGGGACCTTATCTCTTCTTGGTATCGTAAATCTTGTGAATGGCTTGTCTGC	11077
QY	206	PheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHis	225
DB	11078	CATGATGACGCTTCGACCTTT-----CTGCGGACAGCTCTAGCT	11116
QY	226	GlnGlnAlaAspTyrLeu-----SerThrHisTyrAlaGluGlnGlyAsnHisArgLeu	243
DB	11117	GTACAAGCGGACGCGCTGTGCAACAGTAGAATATCATCTCTGGAAATCATCTATTTC	11176
QY	244	PheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAspSerPro	263
DB	11177	GCTAATGGGAAGCGGTGGTGTGTTTCCCGGC---GCTATCTCAGGCCGATGCTGAT	11233
QY	264	ArgTrpArgGlnThrGlyLysSerValLeuAsnThrGluLeuLysGlnValTyrAla	283
DB	11234	CCTTGGCTTGGCAAGGGTTACGGATCTTACGAGAGAGATCTCTCTGAGCAATTTCTGAAT	11293
QY	284	AspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspilePheLeu	303
DB	11294	GATGTGGGCATVTTCCAAATGTCTCCGATGATATCCGACCTTTTACTC-----	11341
QY	304	LysAlaTyrGlySerAlaLysArgValAsnLeuGlu-----	315
DB	11342	-----TGGGATATGTGGACCTGTGATCTCTCCACGCGTCCGGCCTACCTGATTG	11395
QY	316	---LysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuLeu	334


```
Db 11396 GCTGAGCGTACCGCAA---TGCGAGAGTGTGTGTCACAGATGAAGTCTCGT 11452
QY 335 SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLys 354
Db 11453 AGCATGACGATCCGGATGGTAGAATTAGCTTCTCAATGAT----- 11494
QY 355 AsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIle 374
Db 11495 -----GCTGCATTCCGA-----ATTGCACCTGAATATGAAGATATT 11530
QY 375 LysTyrPheAla-----ThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLys 392
Db 11531 GCAGCGTAGCCCAAGCGCTCGATATATCACCTCCAGCTCATGAACACACCTGGCGG 11590
QY 393 AlaLeuSerAsnAla---GlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThr 411
Db 11591 ATTATTAATTTCGGCTACAGTTATGCTGTGTAATCTCTCCGATGCGGTGAAGCTATA 11650
QY 412 ValMetValLeuLysAlaSerPro-----ProGlyGluPheHisAlaGlnProAsp 428
Db 11651 TTGCATCTGCAAGGTGCGTCCGCTACTATCAGCCGGG-----CATGCGCATGCCGAC 11704
QY 429 AsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPhe 448
Db 11705 ACATTGAGTTTGAAGTAAAGCGCTTTGGAAACGGCTTAGTAGTCAATCCGGAATTCT 11764
QY 449 ValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIle 468
Db 11765 CAGTACGGTGAATCACTCTGAG-----CGACAAAGGCGGCTGCTACTGCGGCG 11812
QY 469 HisSerThrLeuThrLeu 474
Db 11813 CATAATACGGTGGGATTG 11830

RESULT 15
AL646082/2
LOCUS
DEFINITION Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
segment 7/11.
ACCESSION AL646082 AL646053
VERSION
KEYWORDS
SOURCE
ORGANISM Ralstonia solanacearum.
Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
1 (bases 1 to 205050)
Salanoubat,M., Genin,S., Artiguenave,F., Gourzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choise,N., Claudel-Renard,C., Cunnac,S., Denange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,I.,
Siguier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
2 (bases 1 to 205050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, IMGM CNRS
118 Route de Marbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Génétique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
Location/Qualifiers
1..205050
source
```

```

/organism="Ralstonia solanacearum"
/strain="GMI1000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
complement(486..845)
/gene="RSP0956"
/notes="RSP0956"
/notes="RSP0956"
complement(486..845)
/gene="RSP0956"
/function="miscellaneous; unknown"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD18107.1"
/db_xref="GI:17431429"
/translation="MSEKLGAERLMENWRDDLEALDLRIVRPAICQVKLLPEPGVIQ
RVLDNDARDCIHGNOTAFASRLGLFLHFQMKDLAEAYGPSDATEIVRRVWAHLOPR
VSDQLGSLPNPGALSAV"
1409..1714
/gene="RSP0957"
/notes="RSP0957"
1409..1714
/gene="RSP0957"
/function="miscellaneous; unknown"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD18108.1"
/db_xref="GI:17431430"
/translation="MEHDKYLHAGISAVIASGVTTLAVDSPNRFWYGLGAGMAVGLTK
EIVDRKTNGRDPSDLLADLAGALIGAYADSLRPAVFKPTGYAYGVQLNVAFD"
complement(1869..2540)
/gene="RSP0958"
/notes="RSP0958"
complement(1869..2540)
/gene="RSP0958"
/function="miscellaneous; hypothetical/global homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD18109.1"
/db_xref="GI:17431431"
/translation="MPLDQPLGLHARNIPGATGIFHEYQLDFCCGQHSLRDAQAQK
GIDAAPIAARLQALQAEATPDGAVDWSAVSPSMLIDHILRPHRERHQLPELIRAR
RVEHVHGRPCPCVGLSLLLEAMQELSHMKQKQILFPMIARGHGLRAGGPIAVMR
MEHDHGGALQRLMTLTNDITPPRAACTTWRALYGLSVFREDLMEHILHNNILFEG
AAVAD"
2706..4238
/gene="RSP0959"
/notes="RSP0959"
2706..4238
/gene="RSP0959"
/function="miscellaneous; not classified regulator"
/notes="Product confidence : probable
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
```

```

predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE SIGMA-54 INTERACTING TRANSCRIPTION
REGULATOR PROTEIN"
/protein_id="CAD18110.1"
/db_xref="GI:17431432"
/translation="MIDRLILLADLVADIPPAVVRMQLRVGSLRSHFRCAVALLRLEED
HLRPAVDGLVDAIGREFVVGQHPRLAAIILARGVTCFHDSTLPIYDGLIDQAG
APLPVHDCMGSVHVEGQWGLATLDTLKIIGTFDAQAELRQIAVVEAAATRLLE
AELRALRATRGDLPDPAADDWETIGQSPAIHOLLHELVVADVSELVLLIGETGVG
KEIFARHLRLRRDRPLVAVWVCAALPESLAESLFGHVKGAFSGVGRPEFAA
DGDTLDEVEGLPLPVQAKLLRALONGEIORLGRDHPRHVDARVIAATNRLERVR
DGAFRADYHLRLSVTPVPIPLRERGNVULLLAGRFLESNRLARGLRLSPGAEHA
LRRYPWGNREHVLVSRAALKMISRGASDDIVTLETGLLDLDAQVAPPSFAGLP
PEASPIAGTTLRDAVADCQRRHIIQAALAIHQDNWAQTARTLGLDASNLHLARRLG
LK"
complement(4338..5294)
/gene="Rsp0960"
/note="RS02291"
complement(4338..5294)
/gene="Rsp0960"
/EC_number="1.-.-.-"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE OXIDOREDUCTASE PROTEIN"
/protein_id="CAD1811.1"
/db_xref="GI:17431433"
/translation="MGARVPKILVVVGANGQIGSELALAEYRHAQVITSLVPAGR
HPHLEMLNATDRCELAVERHGITOIYLLAALSATGEKAPQWMDLNNRSLNV
LEKQMGQVERVPESSIAVGPPTPRDGTPOKTYMETVVGISQAGEGCWVHA
NHGVDRMRYPLGLSYKTLPGGTTIDYADIFHAAVKEPYTCFLDDEALPMYMP
DAIRATILMEAPARLTERGSYNTAGMTFTPAQAAALREHVFGFQIRYEDYRQAI
QAQWPDSDISVARADWQWQAEYGLIESVMADMLVNLRTLLAA"
complement(5294..6493)
/gene="kbl"
/note="Rsp0961; RS02292"
complement(5294..6493)
/gene="kbl"
/EC_number="2.3.1.29"
/function="small molecule metabolism; central intermediary
metabolism"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE 2-AMINO-3-KETOTUTYRATE COENZYME A
LIGASE PROTEIN"
/protein_id="CAD18112.1"
/db_xref="GI:17431434"
/translation="WPNAAFVAFIRTELDAIREAGLFKQERVIAASPOGARVRTIDGR
EYINLCANNYGLSSHPRVDAHEALRTHGFLSSVRFCGTQDLHKLTLRLSAFL
GTEDTLQSAAPDANGGLPETLLGAEDAVISDALNHSIIDGIRLCKARRRYQHNDL
DLRAQLTQARDAARHLLVFTDGVFMDGTIARLDENKSIDCAHGALLIDECRATG
FLVPGRGTHLRVGVKVDIITGLKALGASGGF"SARKEVVALLRQSRPYLFS
NTPVAIIIGASIAVDILEGSTELDRLENRFRFRAGLEALGFVVKGEHPHIIPIMV
YADAKQAQLAQRLLLELVVVGFFVFPVVKQARIRVQMSAVHDEAALQAALGAFQA
CHELGLI"
6665..7213
/gene="Rsp0962"
/note="RS02293"

```

CDS

```

6665..7213
/gene="Rsp0962"
/function="miscellaneous; not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD18113.1"
/db_xref="GI:17431435"
/translation="MGIALQGLRQRLSLDELSRRAGSVKSMLSQISRLNLTNPTVAV
LWRLANALGVSLDTFLQASAGEPAGNITVFPHPALPKSPFTRCDRLIGFDLAG
RFWEYELTIQAGVLSSEPHGEHTEHLSVLSSGNTVTDADKKLRHGETARYRADY
ARHVIANGGKTTATALLVVVHPA"
Alignment Scores:
Pred. No.: 0.283 Length: 205050
Score: 160.50 Matches: 129
Percent Similarity: 35.26% Conservative: 85
Best Local Similarity: 21.25% Mismatches: 274
Query Match: 4.59% Indels: 119
DB: 1 Gaps: 31
US-09-802-285A-2 (1-659) x AL646082 (1-205050)
QY 49 LysAlaValAlaAlaGlyAsnTyrAspAlaAlaLysAlaLeuLeuAla----- 65
Db 66101 AAGCGCGCGCGAGGGGCTGTGTGGCCATCGCTGGAGAGACCGCTCGCGCTAT 66042
QY 66 -----TyrTyrArgGluLysSerLysAlaArgGluProAspPhe---SerAsnAlaGlu 82
Db 66041 TTCGCTGTGTACACCGCGCGAGCAGCGC---GAGCGGAGTGGCACCGCAATCCGTTC 65985
QY 83 LysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAspLysAla 102
Db 65984 AACGGCGCGCGGTCAACAGAGCCC---TCGAGCCATGGTGGCAGATCGCGAT----- 65934
QY 103 LeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLysAspIle 122
Db 65933 -----TTCGACCCCGCGCTGGGC----- 65916
QY 123 AsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArgValLys 142
Db 65915 -----GACATCAAGCGGATTTGGGAGGCGGTTCGCGCTTCGAC 65880
QY 143 TrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla----- 160
Db 65879 TGG-----GTCCTGGTCTTGGCGGAGGAGGGTTCGGCGCGCGCCCGCCATGACT 65826
QY 161 -----ArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeu 178
Db 65825 CAGCTCAATGCGTGGCTG-----ACGTCTCGGCGCGGCGCAATCGCGCTATCTC 65775
QY 179 SerGlnAspAsnAspLysPheValTrpArg---ProLeuGluValSerAspArgValGln 197
Db 65774 GGTCCCAAC-----TGAAGTGGCGGAGAGCGCGGATCTCGCGTCATG 65730
QY 198 SerLeuProProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeu 217
Db 65729 CACCTGGCGATGGCGCGCTGTGTGGGCGAGCACTTCTCGCGCGCGCGCG---CTG 65673
QY 218 MetGluPheLeuAsnSerTyrHisGlnGln---AlaAspTyrLeuSerThrHisTyrAla 236
Db 65672 ATGGCGCTGTCCCGCACCGACCTGCGAGCGCATCGCGCGACCTCGGCTACGCCATCGCC 65613
QY 237 GluGlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPhe 256
Db 65612 CAGGACAATACACCGGACCTCGGAGGCTGCGCG---CTGTTTCATCGCGGAGCTGG 65556

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 20:50:03 ; Search time 39 Seconds
(without alignments)
2006.736 Million cell updates/sec

Title: US-09-802-285A-2

Perfect score: 3494

Sequence: 1 MTKTKFKRIIVFAVIALSSG.....KGLKLTLTLINGKQQLVLP 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Lasting first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128	3.7	25	10	US-09-802-285-3
2	128	3.7	25	15	US-10-291-337-3
3	116	3.3	4563	11	US-09-802-640-32
4	115.5	3.3	1331	10	US-09-801-368-370
5	113.5	3.2	1398	14	US-10-090-624-6
6	110.5	3.2	563	9	US-09-815-242-5665
7	110.5	3.2	578	9	US-09-815-242-12206
8	110	3.1	4563	10	US-09-870-759-128
9	107.5	3.1	1336	11	US-09-934-455-22
10	107.5	3.1	1336	15	US-10-278-173-128
11	106	3.0	1375	9	US-09-740-274-4
12	105.5	3.0	1167	9	US-09-815-242-11522
13	105	3.0	965	11	US-09-842-484A-2
14	105	3.0	965	15	US-10-184-485-3
15	105	3.0	1135	11	US-09-759-130B-42

15	104	3.0	1849	15	US-10-242-056-49	Sequence 49, Appli
17	104	3.0	2516	9	US-09-817-514A-2	Sequence 2, Appli
18	104	3.0	2516	15	US-10-242-056-47	Sequence 47, Appli
19	103.5	3.0	1183	10	US-09-870-759-45	Sequence 45, Appli
20	103.5	3.0	1599	14	US-10-092-880-9	Sequence 9, Appli
21	103	2.9	920	9	US-09-815-242-12995	Sequence 12995, A
22	103	2.9	920	9	US-09-815-242-13148	Sequence 13148, A
23	102	2.9	26926	10	US-09-759-508B-2	Sequence 2, Appli
24	101.5	2.9	943	9	US-09-969-362-5	Sequence 5, Appli
25	101.5	2.9	964	15	US-10-100-049-74	Sequence 74, Appli
26	101.5	2.9	965	11	US-09-842-484A-4	Sequence 4, Appli
27	101.5	2.9	965	15	US-10-217-613-3	Sequence 3, Appli
28	101.5	2.9	965	15	US-10-217-613-8	Sequence 8, Appli
29	101	2.9	917	9	US-09-815-242-5503	Sequence 5603, Ap
30	101	2.9	920	9	US-09-815-242-12181	Sequence 12181, A
31	101	2.9	972	15	US-10-217-613-9	Sequence 9, Appli
32	100.5	2.9	664	9	US-09-815-242-12577	Sequence 12577, A
33	100.5	2.9	691	9	US-09-815-242-5223	Sequence 5223, Ap
34	99.5	2.8	690	9	US-09-815-242-12460	Sequence 12460, A
35	99.5	2.8	703	9	US-09-815-242-5610	Sequence 5610, Ap
36	99.5	2.8	804	14	US-10-011-588-37	Sequence 37, Appli
37	99	2.8	633	9	US-09-871-615-5	Sequence 5, Appli
38	99	2.8	1477	14	US-10-092-880-4	Sequence 4, Appli
39	98.5	2.8	556	9	US-09-815-242-5527	Sequence 5527, Ap
40	98.5	2.8	565	9	US-09-815-242-12352	Sequence 12352, A
41	98.5	2.8	565	9	US-09-815-242-12931	Sequence 12931, A
42	98	2.8	615	10	US-09-908-193-26	Sequence 26, Appli
43	98	2.8	944	15	US-10-213-990-27	Sequence 27, Appli
44	98	2.8	1187	9	US-09-768-436-4	Sequence 4, Appli
45	97.5	2.8	515	9	US-09-815-242-4928	Sequence 4928, Ap

ALIGNMENTS

RESULT 1

US-09-802-285-3
; Sequence 3, Application US/09802285
; Patent No. US20020122793A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Dongfang
; APPLICANT: Pojasek, Kevin
; APPLICANT: Shriner, Zachary
; APPLICANT: Holley, Kristine
; APPLICANT: El-Shabrawi, Yusuf
; APPLICANT: Venkataraman, Ganesh
; APPLICANT: Sasisekharan, Ram
; TITLE OF INVENTION: Heparinase III and Uses Thereof
; FILE REFERENCE: M0656/7063HCL
; CURRENT APPLICATION NUMBER: US/09/802,285
; CURRENT FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Flavobacterium heparinum
US-09-802-285-3

Query Match 3.7%; Score 128; DB 10; Length 25;
Best Local Similarity 100.0%; Pred. No. 3.8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 QYADGMQFELSPIYHVAIDFLK 304

Db 1 QYADGMQFELSPIYHVAIDFLK 25

RESULT 2

US-10-291-337-3
; Sequence 3, Application US/10291337
; Publication No. US2003009628A1
; GENERAL INFORMATION:

```
; APPLICANT: Liu, Dongfang
; APPLICANT: Pojasek, Kevin
; APPLICANT: Shriver, Zachary
; APPLICANT: Holley, Kristine
; APPLICANT: El-Shabrawi, Yusuf
; APPLICANT: Venkataraman, Ganesh
; APPLICANT: Sasisekharan, Ram
; TITLE OF INVENTION: Heparinase III and Uses Thereof
; FILE REFERENCE: M0656/7063HCL
; CURRENT APPLICATION NUMBER: US/10/291,337
; CURRENT FILING DATE: 2002-11-08
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Flavobacterium heparinum
US-10-291-337-3

Query Match
Best Local Similarity 100.0%; Pred. No. 3.8e-05; Length 25;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 QYADGMQFELSPIYHVAIDIFLK 304
DB 1 QYADGMQFELSPIYHVAIDIFLK 25

RESULT 3
US-09-802-640-32
; Sequence 32, Application US/09802640
; Publication No. US20030036057A1
; GENERAL INFORMATION:
; APPLICANT: Braun, Andreas
; APPLICANT: Bonsal Aruna
; APPLICANT: Kieyn Patrick
; TITLE OF INVENTION: GENES AND POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE AND THEIR USE
; FILE REFERENCE: 24736-2048
; CURRENT APPLICATION NUMBER: US/09/802,640
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 4563
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-802-640-32

Query Match
Best Local Similarity 16.3%; Pred. No. 1.7; Length 4563;
Matches 104; Conservative 94; Mismatches 197; Indels 244; Gaps 26;

QY 218 MEFLNSY-----HQADY-----LSTHYAEGCNHRLFEA-----QRNL----- 250
DB 3068 IDFLNVALFLSPSAQAQSWQSAFNFQYKYNQNFSAAGNNENIMEAHVINGEANLDFLN 3127
QY 251 -----FAGVSPFEKDSPEWROTG----- 269
DB 3128 IPLTIPENRLPYTITTPPLKDFSLMEKTLGKLEFKLTKTKQSFDSLVSVAQYKKNKHHSIT 3187
QY 270 --ISVLNTEIKQKVADGMQFELSPIYHVAIDIFLKAYGSAK-----RVNLEK-----BFPQ 320
DB 3188 NPLAVLCBFIQSQISKSPDRHFENK-----RNNALDFVTKSYNETKIKFDYKAEKSHDELPR 3244
QY 321 SYVQTVENMVALISISLPDNTPMFGDSWITDKNFRMAQF-----ASWARVFPANQATKY 376
DB 3245 TF-----QIEPGTVPVV-----NVEVSPTEIEMSAFGYVFPKAVSMPS 3282
QY 377 FATDCKGQKAPNF-----LSKALSNAGF-YT 401
DB 3283 FSLIGSDVRVPSYTLILPSLELPLVLHVPRLKLSLPHEKELCTISHIFIPAMGNITYDFS 3342

; APPLICANT: VMLKASPPGFHQAQPDNGTPELFITKGRNFTPDAGVFVYSGDEAL----- 456
; APPLICANT: VITLNTNAELFNQSD--IVAHLLSSSSSVIDALQYKLEGTTRTRKRG 3392
; APPLICANT: MKLRNRYQTRIHTLTDNQNMVITKARQNKWE-----TGNLNDVL 498
; APPLICANT: LKATALSLSNKFVEGSHNSTVSLTTKMEVSVAKTTKAEIPIILRMNFQKQLNGTKSKP 3452
; APPLICANT: TYTNPSPYNDHQHRSVLFINK-----YFLVIDRAIGEATGNI----- 536
; APPLICANT: TVSSMEFKYDFNSMLYSTAKGAVDHKLSLESLTSYFIESSTKGDVKGSVLSREYSGT 3512
; APPLICANT: GVVH-----WOLKEDSNPVEFKTKRVRVTVYRDG--NNLM 568
; APPLICANT: IASEANTYLNKSTRESSVKLQTSKIDDIWNLEVENFAGATLORIYLSLHSTKNHLQ 3572
; APPLICANT: IQSL---NADRTS---LNEEGKYSVYVYNKELKPAFVFEKP-----KKNAGTON--- 612
; APPLICANT: LEGLFFTTNGEHTSKATLELSPWQMSALVQVHASQPSFPHDFDLQGEVALNANTKNQKIR 3632
; APPLICANT: FVSIVYPYDQKAPFISIRENKGNDFFKGLNLTLITING 651
; APPLICANT: WKNEVRHSGSFQSQVEL-----SNDQEKAHLDIAGSLEG 3667

RESULT 4
US-09-801-368-370
; Sequence 370, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 370
; LENGTH: 1331
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-370

Query Match
Best Local Similarity 18.5%; Pred. No. 0.28; Length 1331;
Matches 157; Conservative 108; Mismatches 287; Indels 297; Gaps 45;

QY 1 MTTKIFKEIIVFAVIALSSGNILAQS-----SSITRKDFDHINLEYSGLKVNKAV 51
DB 25 ISSKLVKRLSHSHKLSRDLKALGSETISDGPSTFKD-RYVFNESLYLKKLKKTA 83
QY 52 AAGNYDDAAKALLAYREKSKAREPDFSNAEKPADIRQPIDKVTREMAKALVHQFQPHK 111
DB 84 LDDYVTRGKILNRYEEDGDDEIIRLSNG-----DRIDEDL-----HS 122
QY 112 GYGTFD---YKQDINQWQWPKVDNEVRWQ---LHRVKWQAMALYVH---ATGDEKYARE 162
```

123 GVKFFSTTPYCKMR-----SDSDELAWEIATERFKWQSMRLARVLKGDIVKGEKTRIAN 177
QY 163 WYQYSDWARKNPLGLSQNDKFWVRPLE-----VSDRVQSL-----PPTFSLFVNSPA 211
Db 178 QV-----KKP-GLINKELSDIEIWLKAWLNGRTMQEMQSLTYLRDSSDSVFEIMK 228
QY 212 F-----TPAFLMEFLNSYHQOQADY---LSTHYAEQ--GNHRLFEAQ-----247
Db 229 FOIPOGKILSDALEAILQDLNRRHVSVYWPNLKMKYKPIINTAFTARIDVMSW 288
QY 248 -----RNLFAGVSF-----PEFKDSPRWRTGTSVLNTEIKKQVADG 285
Db 289 LNFKNLTLRRQELDDWINRFSPISSSONCQEDFGVPOW-NCKMKILAEQLMKEKNIES 347
QY 286 MQFELSPIYHVAADIFLKAYGSAKRVNLEK-----EFP-----Q 320
Db 348 I-FQKKIEFYPLSPWFKLHLFIYRETLTKMKIKYPYERLRLSLLAFVYLKIVELTRL 406
QY 321 SYVQTVENMIMALISISLPDYNT-----PMFGDSWITDK 354
Db 407 SYARKLKPTMMIDQIDDFAFIRLSVLQKYTLTKYCSNLPFDVDEPTFENTVI---463
QY 355 NFRMAQFASWARVFPANQAIKY-----FATGKQG-KAPNFLSK---ALSNAQFYTF 402
Db 464 -----EAIYLFLLNLKLISSKQNFKAPDLLLKYWDHLKNTGHYI- 505
QY 403 RSGWKNATVMVLKASPPGEFAQDNGTFEL---FIKGRNTPDAGVYVSGDEAIMKL 459
Db 506 -----NGAETVI-----PNEFLKTLRLVHLKQFYLLKQONPPTFA---NASEA---547
QY 450 RNWYQTRHSTLTLDNQMWITKARONKWTGNLQVLTYPNSYPNLHQRSVLFINK 519
Db 548 EKN-----LSSIFENLGMKRLNRF---SNILVAFQNSAVYQINHAQVIL---K 592
QY 520 K-----YFLIVDRAIGATGNLGVHW-----QLKEDSNVPVFKTKNR-----VYTYRDG 564
Db 593 KLDKADHYFLVYS---GNTPESSGVYFAAPPELLGCDNDITLRLNRKNSIGCDLVKPLDYG 649
QY 565 NNLMTQSLNADRTSLN-----BEGKVSY---VYNKELKRFAPVFEKPKKNAQTQNF--613
Db 650 NNLANYDITTKETDLNLSKGEDSKGIPYRVWANSSSDLDRAHQSKKKNFSDPDQ 709
QY 614 -----VSIVVP-----YDGQKAPESIRENK-----GNDEFEKGL- 643
Db 710 HLDKKNVEFEVALSSLCALVLYPGEVWVG---PVYKLPGNLNFASNEMDLKIG 766
QY 644 --NLTLTIN 650
Db 767 NENTLILN 775

RESULT 5

US-10-090-624-6
; Sequence 6, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 1398

; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; US-10-090-624-6
Query Match 3.2%; Score 113.5; DB 14; Length 1398;
Best Local Similarity 18.5%; Pred. No. 0.47;
Matches 118; Conservative 83; Mismatches 233; Indels 203; Gaps 31;
QY 134 VRWQLHRVKKWQAMALVYHATGDEKYAREWVYQYSDWARKNPLG--LSQNDKFWVRPLE 191
Db 703 VEWHIKYV-----GDTEYRTFEIYATEPWIKPFGVSGSVILENTEFV---LR 746
QY 192 VSDRVQSPPTF---SLFVNSPAFTPAFLMELNS-----YHQOADYLSHY---AQ 238
Db 747 VKYDVEGLEPGLYVGRITIDDDPT-TPVIEDEILNTIVPEKFTPENNYTLTWYDINGPEM 805
QY 239 GNHRLFEAQRLNLFAGVSPFEKDSPRWRTGTSVLNTEIK--QVYVADGMQ-----287
Db 806 VTHHFTTPEGVDVLYAMTTYWDYGLYRPGDMFVFPYQLDYLPAAVSNKPMPCGNWELVWVG 865
QY 288 FELSPIYHVAADIFLKAYG-----SAKVNLEKEFPQSYVQTVENMI---MALIS 335
Db 866 FNFAPLYESG---FLVRIYGVETPSVMYINRTYLDNTFEFIEFNITNIYAPINATLIP 922
QY 336 ISLPDYNTPM--FGDSWITDKNFRMAQFASWARVFPANQAIKYFATDGKQKAPNFLSKA 393
Db 923 IGLTYNASVSVSGOGEFFIKGIEVPEGTAEULKIRIGNPSV-----PN-----965
QY 394 LSNAGFYFRSGWKNATVMVLKASPPGEFHA---QPONGTPELFIKGRNFTPDAGVFVY 450
Db 966 -SDLDLYL---DSKGNLVALDGNPTAEVVVVEPKGVYSIVVHGYSVVDENGNTT 1020
QY 451 SGDEAIMKLRNRYOTRIHSTLTLDNQMWITKARONKWTGNLQVLTYPNSYPNLHD 510
Db 1021 TTFDLVWQ-----MTLDNGNIKLDK---DSIILGSNBSVVVTANITTI--DRDH 1063
QY 511 QRSV-----LFINKY-----FLVIDRA-----IGEATGNLGVHWQLKEDSN 547
Db 1064 PFGVSGIIEIRDNEVYQDNTSIAKIPITLVIDRADFAVGLTPAEGVLG-----1113
QY 548 PVFDTKRVVT-----TYRDGNLMTQSLNADRTSLNBEKGVSVYV-----590
Db 1114 -----EARNYTLIVKHALTLEPVENATV--IIGNVYTLTDENGTVTFYAPTGLGSDRI 1165
QY 591 -----NKELKRPAPVFEKPKKNAQTQNFVSVIYPDG--622
Db 1166 TVIVKKENFTLEKTFQITVSEPEITBEDINEPKLAMSPEANA---TIVSVEMSEGGV 1222
QY 623 OKAPEISIRENKGNDFEKGKLNLTITNGKQQLVLP 659
Db 1223 KATVVEITIN-----GTANETATI-----VVPVP 1247

RESULT 6

US-09-815-242-5665
; Sequence 5665, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

```
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5665
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5665

Query Match      3.2%; Score 110.5; DB 9; Length 563;
Best Local Similarity 18.1%; Pred. No. 0.21;
Matches 105; Conservative 94; Mismatches 211; Indels 171; Gaps 26;

QY 91 IDKVTREMAKALVHQFQPHKGYGYDYGKDINQMWPVKDNEVRWQLHRVKKWQAMA-L 149
Db 8 IERILNEFRENLVQVPPQN-----HWSIQLSLETLTGQFAYWSAMGHH 54

QY 150 VYHATG---DKYAREWVYQSDWARKNPLGLSDNKFVWRPLEVSRVOSLPPTFSLF 206
Db 55 MYHSEVWLIDGSKSLTYKEAIAIRILQHMAQSDNQTAVQQHM----- 98

QY 207 VNSPAFTAPFLMEFLNSVHQADYLSTHYAEQGNHRLFEAQORNLFAGVSE-PEEKDSPRW 265
Db 99 -----AQIMSDIDNSIHRTRYLQSNITIDYAEADRYIVSQSLYLGHFPHPTPKSASGF 151

QY 266 RQTGIS-----VLNTEIKQVYADQMOPFELSPI-YHVAADID-----FL 303
Db 152 SEADLEKVAPECHTSFQLHYLAVHQDVLLTRYVEGKEDQVEKVLQYQLADIDI SEIPKDFI 211

QY 304 KAYSARKVNLEKEFPQSYQVTENMIMALLISISLDPYNTMPFGDS----- 349
Db 212 LLPHPYQINVLQHQHPQ-YLQYSEQGLIKGLGVS-----GDSVYPTSSVTRVFSKA 261

QY 350 -----WITDKNF-RMAQFASWARVFPANQAI-----KYFATDGKOGKAP 387
Db 262 LNIYKLPPIHVKIINFRTNDLEQIERIDIAQVIASVKDEVEPHFKLMFEEGYRALLP 321

QY 388 NFLSKA-----LSNAGFYTFRSG-----WDKNATVMVLKASPPGFHQAQDNGTPEL- 434
Db 322 NPLGQTVPEPMDLLTNSAMIV-REGIPNYHADKDIHVLA-----SLFETWPDSPSLS 374

QY 435 -FIKGRNFTPDAGVVFVSGDEAIMKLRNWRQTRI-----HSTLTL-----D 475
Db 375 QVIEQSGLAPEAWLECYL-DRTLPLILKLYSNTGISLEAHVQNSLIELKDGIPDVCVFRD 433

QY 476 NONNVITKARQNKWETGNLDVLTNPSPYENLHQRSVLFINKYFLVIDR-----AI 529
Db 434 LEGICLSRTIATEKQLVFN--VVAASSP-----VVAHDEAHRLKYYVNVNHLGHLVSTI 487

QY 530 GEATGNLGVHWQL-----KE-DSNPVFDKTKNRVYTT 560
Db 488 GKATKNEVVLVQLVAHRLMTWKKEYANNAVDCVEDLYQT 528

RESULT 7
US-09-815-242-12206
; Sequence 12206, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
```

```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12206
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12206

Query Match      3.2%; Score 110.5; DB 9; Length 578;
Best Local Similarity 18.1%; Pred. No. 0.22;
Matches 105; Conservative 94; Mismatches 211; Indels 171; Gaps 26;

QY 91 IDKVTREMAKALVHQFQPHKGYGYDYGKDINQMWPVKDNEVRWQLHRVKKWQAMA-L 149
Db 15 IERILNEFRENLVQVPPQN-----HWSIQLSLETLTGQFAYWSAMGHH 61

QY 150 VYHATG---DKYAREWVYQSDWARKNPLGLSDNKFVWRPLEVSRVOSLPPTFSLF 206
Db 62 MYHSEVWLIDGSKSLTYKEAIAIRILQHMAQSDNQTAVQQHM----- 105

QY 207 VNSPAFTAPFLMEFLNSVHQADYLSTHYAEQGNHRLFEAQORNLFAGVSE-PEEKDSPRW 265
Db 106 -----AQIMSDIDNSIHRTRYLQSNITIDYAEADRYIVSQSLYLGHFPHPTPKSASGF 158

QY 266 RQTGIS-----VLNTEIKQVYADQMOPFELSPI-YHVAADID-----FL 303
Db 159 SEADLEKVAPECHTSFQLHYLAVHQDVLLTRYVEGKEDQVEKVLQYQLADIDI SEIPKDFI 218

QY 304 KAYSARKVNLEKEFPQSYQVTENMIMALLISISLDPYNTMPFGDS----- 349
Db 219 LLPHPYQINVLQHQHPQ-YLQYSEQGLIKGLGVS-----GDSVYPTSSVTRVFSKA 268

QY 350 -----WITDKNF-RMAQFASWARVFPANQAI-----KYFATDGKOGKAP 387
Db 269 LNIYKLPPIHVKIINFRTNDLEQIERIDIAQVIASVKDEVEPHFKLMFEEGYRALLP 328

QY 388 NFLSKA-----LSNAGFYTFRSG-----WDKNATVMVLKASPPGFHQAQDNGTPEL- 434
Db 329 NPLGQTVPEPMDLLTNSAMIV-REGIPNYHADKDIHVLA-----SLFETWPDSPSLS 381

QY 435 -FIKGRNFTPDAGVVFVSGDEAIMKLRNWRQTRI-----HSTLTL-----D 475
Db 382 QVIEQSGLAPEAWLECYL-DRTLPLILKLYSNTGISLEAHVQNSLIELKDGIPDVCVFRD 440

QY 476 NONNVITKARQNKWETGNLDVLTNPSPYENLHQRSVLFINKYFLVIDR-----AI 529
Db 441 LEGICLSRTIATEKQLVFN--VVAASSP-----VVAHDEAHRLKYYVNVNHLGHLVSTI 494

QY 530 GEATGNLGVHWQL-----KE-DSNPVFDKTKNRVYTT 560
```


Db 724 NPILL-----TDPMEDQYPPPSDTLLKYV-----SEILMEESNGDYKQSM 763
QY 468 IHSTLT-----DNQNMVITKARO---NKWETGNLVLTYTNSYNLHDOR 512
Db 764 FYDSLALRKTEMLQOVITDSQNSQSFSPADSLITNSWDASGIDBSAYSADPQPVNEIMV 823
QY 513 SVLFINKKYFLVIDRAIGEAATGNL---GVHWOLKEDSNFVFKTKNRVVTYRDGNNLMIQ 570
Db 824 KSMFSDAESALQFKKGVEASKEFLPNSQWVNLND-----IERSEKRSVKEEMGLDQLRV 879
QY 571 SLNADRTSLNEEGKVSIVYNNKELKPAFVFEKPKKNAGTQNFVSVVYPYDQKAPET-- 628
Db 880 KKNHER--DFEEVRSSKQFASNVEDSKVTDMDFK-----VLLLDGECDDPOTLL 925
QY 629 -----SIRENKGNDFEKGK 642
Db 926 DSEIQAIRSSK-NIGERK 943

RESULT 10

US-10-278-173-128
; Sequence 128 Application US/10278173
; Publication No. US20030061637A1

; GENERAL INFORMATION:

; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Pineda, Onaira
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Keddle, James
; APPLICANT: Heard, Jacqueline
; APPLICANT: Reuber, Lynne
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond

; TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION

; FILE REFERENCE: MBI-009

; CURRENT APPLICATION NUMBER: US/10/278,173

; CURRENT FILING DATE: 2002-10-21

; PRIOR APPLICATION NUMBER: US/09/533,392

; PRIOR FILING DATE: 2000-03-22

; PRIOR APPLICATION NUMBER: 60/125,814

; PRIOR FILING DATE: 1999-03-23

; NUMBER OF SEQ ID NOS: 177

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 128

; LENGTH: 1336

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; OTHER INFORMATION: G987

US-10-278-173-128

Query Match

Best Local Similarity 3.1%; Score 107.5; DB 15; Length 1336;

Matches 136; Conservative 110; Mismatches 252; Indels 241; Gaps 39;

QY 24 AQSSSIT-RKDFDHINLEYSGLKVNKAVAGNVDAAKALLAYREKSKAREPDDF----- 78

Db 326 AQAVSINDRTADEL-----LSRIQ--HSSYGDGTERLAHYFANSLARLAGIGTQV 377

QY 79 -----SNAEKPADIRQ-----PIDKVTREMAKALVH--QFQPHKGYGYFDYGDKI 122

Db 378 YTALSSKKTSTSDMLKAYCTIVISVCPFKKIIIFANHSIMRLASSANAKTHIIDFGISD 437

QY 123 NQWMPVKDNVNRWOLHVRKVMQAMALVYHTGDE-----KYARE 162

Db 438 GFQ-WPS-----LIHRLAWRSGSKLRITGIELPQGRPAEGVETGRRLAKYQCK 489

QY 163 W--VYQYSDWAEKNPLGLSQNDKFWRPLEVSD---RVQSLPPTFSLF----- 206

Db 490 FNIPFEYNIAIQK-----WESIKLEDLKLKEGEFVAVNSLFRFRNLDETVA 536
QY 207 VNSP-----AFTPAFLMEPLNSYHQADYLSLTHYAEQGNHRLFEAQRNLPA 252
Db 537 VHSRPRDTVLKIRKIPDVFIPIGLS---GSYN--APPFVTRPREV-----LFH 580
QY 253 GVSPEPFKDSRWRQTGISVLNTBIKKQVYADGMQFELSPLHYHVAADIFLKYAGSAKRV 312
Db 581 YSSLFDMCDTNTLTREDPMRWV---FEKEFYG-----REIMNVVACE-----GTERV 623
QY 313 NLEKEFPQSYVQ-----TVENMIMALISLPLDYNTPMFGDSWITKCNFRMA 359
Db 624 ---ERPESYKQWQARAMRAGFRQIPLKELVOKLKMV-----ESGYKPKREFVD 670
QY 360 QFASWAR-----VFPANQAIKYFATDGKQKA---PNFLSKALSNAAGFYTFRSGWDK 408
Db 671 QDCHWLQGWKGRIVYGSSIWPPFFVYVGRATRVLIMDPNF--SSSLN--GREYP---DG 723
QY 409 NATVMVLKASPPGEFHAQPDNGTPELFITKGRNFTPDAGVFVYSGDEAIMKLNNW--YRQTR 467
Db 724 NPILL-----TDPMEDQYPPPSDTLLKYV-----SEILMEESNGDYKQSM 763
QY 468 IHSTLT-----DNQNMVITKARO---NKWETGNLVLTYTNSYNLHDOR 512
Db 764 FYDSLALRKTEMLQOVITDSQNSQSFSPADSLITNSWDASGIDBSAYSADPQPVNEIMV 823
QY 513 SVLFINKKYFLVIDRAIGEAATGNL---GVHWOLKEDSNFVFKTKNRVVTYRDGNNLMIQ 570
Db 824 KSMFSDAESALQFKKGVEASKEFLPNSQWVNLND-----IERSEKRSVKEEMGLDQLRV 879
QY 571 SLNADRTSLNEEGKVSIVYNNKELKPAFVFEKPKKNAGTQNFVSVVYPYDQKAPET-- 628
Db 880 KKNHER--DFEEVRSSKQFASNVEDSKVTDMDFK-----VLLLDGECDDPOTLL 925
QY 629 -----SIRENKGNDFEKGK 642
Db 926 DSEIQAIRSSK-NIGERK 943

RESULT 11

US-09-740-274-4

; Sequence 4, Application US/09740274

; Patent No. US20020031826A1

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Glucan-containing Compositions and Paper

; FILE REFERENCE: 0357CRD

; CURRENT APPLICATION NUMBER: US/09/740,274

; CURRENT FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: 09/210,361

; PRIOR FILING DATE: 1998-12-11

; PRIOR APPLICATION NUMBER: 09/007,999

; PRIOR FILING DATE: 1998-01-16

; PRIOR APPLICATION NUMBER: 08/478,704

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: 09/009,620

; PRIOR FILING DATE: 1998-01-20

; PRIOR APPLICATION NUMBER: 08/485,243

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: 09/008,172

; PRIOR FILING DATE: 1998-01-16

; PRIOR APPLICATION NUMBER: 08/482,711

; PRIOR FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 4

; LENGTH: 1375

; TYPE: PRT

; ORGANISM: streptococcus mutans

US-09-740-274-4

Query Match 3.0%; Score 106; DB 9; Length 1375;

Best Local Similarity 18.5%; Pred. No. 2.3;

Matches	147;	Conservative	89;	Mismatches	252;	Indels	306;	Gaps	41;
Qy	16	ALSSGNILAOSSITRKDFD		----	HINLEYSGLEKVNKAAGNYDDAAKALL	64			
Db	641	ALSVALLTNKSVPVVYGMFTDDGQYMAHKTINVEAIETL		----	----	688			
Qy	65	AYYEKSKAREPDSNAEKPADIR		----	OPIDKVTR	99			
Db	689	KYVSGGQAMRNOQVGNSEIITSVRYGKGALKATDTGDRTRTTSVGAVIEGNNPRLRLKAS		----	----	748			
Qy	100	DKALVHOFQPHKGKGY		----	FDYKGDINNQMWPVK	145			
Db	749	DRVVNNGAAHKNQAYRPLLLTTDNG		----	IKAYHSDQEAAGLVRYTNDGE	796			
Qy	146	AMALVYHATGDEKYAREWVQY		----	SDWARKNPGLGSQDNDFVWRPLEVSRVQSLPPTFS	203			
Db	797	---LIFTAADIKGYANPQVSGYLGVW		----	PVGAAADQD	842			
Qy	205	LFVNSPAPTFAPLMEFLNSVHQQADYLSVTHYAEQG		----	NHRLFEAQRNLFAGVSPFEPKDS	262			
Db	843	-----KSVHQN		----	ALDSRVMEFGFSNFOAFATKBEETNNVIA	882			
Qy	263	PRWRQTGISVLNTEIKKQVYADGQMFELSPIYHVAAD		----	IFLKAYGSAKRVNLEK	316			
Db	883	DKPAEWGVT		----	DFEMAPQY	926			
Qy	317	EPFQSYQVTVENMIMALLISLDPDYNTPMGDSWITDKNFMAQFASWARVPFANQAIKY		----	----	376			
Db	927	SKENKY		----	GTADDDLVKAT	978			
Qy	377	-----FATDGK		----	QKGAEN	404			
Db	979	GTFVAGSQIKNTLVVVDGKSSGKQQAQYGGAFLEELQAKYFELFARKQIS		----	----	1030			
Qy	405	GWDKNATVMVLKASPP		----	GEFHAQPDNGTTFEL	443			
Db	1031	GVPMDFPSVKIKQWSAKYFNGTNGILGRGAGYVLKQDQATNVFSLVSDNTFLPKSLVNPENHG		----	----	1090			
Qy	444	-----DAGVFVY		----	SGDEA	484			
Db	1091	TSSSVTGLVDGKGYVYTSNGQAKNAFISLGNWY		----	----	1141			
Qy	485	-----RONKETGNLNDLVLTYNTPSYENLDHQRSVLFINKKYFLVIDRAI		----	----	529			
Db	1142	INGANYFLSNGIQLRNAIYDNGNK		----	VLST	1184			
Qy	530	GEATGNLGVHWQIKEDSNPNVPDKTKRNVYTYTRDGNMIMTQSLNADRTSLNEBEGKVSIV		----	----	589			
Db	1185	-----FGQWRYFQNGIMAVGTRVHCAGVQYFDAS		----	GFQAKGQFITTADGKLRY	1232			
Qy	590	YNKELKRPAPVFEKPKKNAGTQNFVSIYVPYDQKAPESIRENKG		----	----	645			
Db	1233	-----FDRSDGNQISNRFV		----	-----RNSKGWFLFDHNGVAVTG	1265			
Qy	646	TLTLNKGQQLVLVLP		----	659				
Db	1266	TVTFNG		----	QRLYFKP	1278			

RESULT 12

```

US-09-815-242-11522
; Sequence 11522, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of

```

```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows version 4.0
; SEQ ID NO 11522
; LENGTH: 1167
; TYPE: PR1
; ORGANISM: Helicobacter pylori
; US-09-815-242-11522

```

Query Match	3.0%	Score 105.5;	DB 9;	Length 1167;
Best Local Similarity	19.1%;	Pred. No. 1.9;		
Matches 150;	Conservative 107;	Mismatches 298;	Indels 231;	Gaps 35

Qy	3	TKIFKRIIVFAVIALSGNII	-----AOSSSITRKDF-----	DHINLEYSGLEKV	47
Db	45	TFWSQRAISFCSTQTSKNIKDSFETIMCEYDEELKKXSFKNLKISIDHVDGTMCNKERL	104		
Qy	48	NKVAAGNYDDAAKALLAYREKSKAREPD	-----FSNAEKPADIQPIDKVTREMA	100	
Db	105	DKLENLNTFFPNICKVLSNARCLSEGVDVPALDSVIFFDGSRAMVDIIQAVGRVNRKAKN	164		
Qy	101	KALVHQFPQPHKGYYFDYCGDKINQWMPVKONEVRWQLHRVK	-----WQQAMALVHYA	153	
Db	165	K-----KXGYIILPIA-----	LRSEIKNLDEAVKNTNFQNIWKVLKALRSHD	207	
Qy	154	TG--DEKYAREWVVOYQDWARKNPLGLSODNDKFWRRELEVSVDROQSILPPTFSLFVNSP	210		
Db	208	SSLVDEAIFKEKIKIFGSDDASN	-----DDEELQKDKTEOSS-----	NDBK	250
Qy	211	-----ATTPAFLEFLNSYHQ-----	QADYLSTHYAEQGN--HRLFEAORNLFEAGVSF	256	
Db	251	EAQKTLFDAILLQDLANAVNMVMTKLGDRNYWENFAKKTGNIARTINERLKELFG--KN	308		
Qy	257	PEPKDSPRWQTGISVLNTEIKKVOYADGMOFELSPIVHVAAIDIFLKAYGSAKRVNLEK	316		
Db	309	PEIFDNFLTSLRG--NIHQSIKEEELDMI-----	ISHIITKPIFDAIFGD-----	NIKN	356
Qy	317	EFQSQSVQTVENMIMALISLSPDYNTPMFGDSMTDKNFRMAQFASWARVPPANQAIKY	376		
Db	357	PI-----AKADLKMVLKJSDJLEGETKDL-----	KNL-----	YESVKT	390
Qy	377	FATDGKGKAPNFLSKALSNAFG--YTFRSGWDK-----	NATVMVLKASPP	420	
Db	391	EAARAKSQSQOELIKNLNTFFKEAFRQSEKLGIVYPTIEVDVDFILRATNGILKKHEN	450		
Qy	421	GEFHAQ-----PDNGT-----	FELFIKGRNFTPD-----	AGVFVYSGDEAIMKLRNW	462
Db	451	TDPNDSQITTFDPFMGTGSPFIARLLSKENDFISDEALKEFKQGLFAPD-----	IVILSY	505	
Qy	463	YRQTRIHSITLTDNONNVITKARONKWEITGN	-----LDVLYTNTPS--	YENLDHQ	511
Db	506	Y-----IALIN-----	ITQAAQRDSSLKFNKXIALTDSLQYLEKNDKGVIPGFBI	553	
Qy	512	RSVLFINKKYFLVIDR-----	AIGEATNGLGVHMQLKEDSNPVPDKTKNRVYTTYTYRDGNN	566	
Db	554	FEDLKENKEIKTWEKONIRVIIGNPPYSGAKSENDDNONLHPKLEKRYVYTY--GKN	611		

Qy 567 LMIQSLNADR-----TSLNEBEGKVSYYVYNKEL--KRPAFVF----- 601
Db 612 STAQNKSTRDTLHISIRMASDLDKRGVLGVVNGSFDISKSDGFRKCAVQDFSHLYA 671
Qy 602 -----EKPK-----NAGTQNFYSIVVPYDQKAPETISIRENKGNDEPKGLN 644
Db 672 LNLGNARTSGEERKKGDDGFDGSRVTVALIFFVVKDAPNHTIYYEVEDYLKREAK 731
Qy 645 LTLTIN 650
Db 732 LNLLAN 737

RESULT 13
US-09-842-484A-2
; Sequence 2, Application US/09842484A
; Publication No. US20030104601A1
; GENERAL INFORMATION:
; APPLICANT: DEANGELIS, PAUL L.
; TITLE OF INVENTION: CHONDROITIN SYNTHASE GENE AND METHODS OF MAKING AND USING SAME
; FILE REFERENCE: 4605.003
; CURRENT APPLICATION NUMBER: US/09/842.484A
; CURRENT FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/199,538
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 965
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-09-842-484A-2

Query Match 3.0%; Score 105; DB 11; Length 965;
Best Local Similarity 17.7%; Pred. No. 1.6;
Matches 131; Conservative 115; Mismatches 238; Indels 258; Gaps 38;

Qy 13 AVIALSSGNILAQSSITRKDF-----DHIN-----LEYSGLEKVNKAVAAG 54
Db 95 SITGKKSNAEIRKVELVPKDFPKDLVLAFLPDHVNDFTWYKNRKSIGIKVPVKNIGLS 154
Qy 55 -----NY-----DAAKALLAYYREKSKAREPDFSNAEK 83
Db 155 IIIPTFNRSRLDITLACLNVQKNTYFVVVAVDGGSKENLLTIVQK-----YEQ 204
Qy 84 PADIRQPIDKVTREMAKALVHQPQPHKGYG-----FDYKGDINQWMP- 128
Db 205 KLDIK-----YVRQKDYGYQLCAVRNLGLRTAKYDFVSIILCDMAPO 246
Qy 129 -----VKDNEV-----RWQLHRVKKWQAMALVYHATGDEKYAREWYQYSDW 170
Db 247 QLWVHSYLTLELNDIVLIGPRKYVDTHNTAQFL-----NDPYLIESLPETA-- 296
Qy 171 ARKNPLGLSQDNKDFVWR--PLEVSDRVQSLPPTFSLFVNSPAPTPAFLEFNLNS--YH 225
Db 297 TNNPSTISKGNISLDWRLEHFKTDNLRLCDSPPRYFV---AGNVAFSKEWLNKVGWFD 353
Qy 226 QADYLSTHYAEOGNHRLFEAQ---RNLFAGVSP---PEFKDSPRWQTGISVLNTEIKK 279
Db 413 KVPYIYRKLPIEDSHIRIPLVSIYPAYNCA-----NYIQ---RCVDSALNQ 458
Qy 280 QV---YADGMQFELSPIYHVAADIFLKAYGSAKRVNLEKFFPOSYYQTVENMMALISI 336
Db 337 SLDPYNTPMFGDSWITDKNFRMAQFASWARVFPANQAIKYFATDGKQKGAENFLSKALSN 396
Qy 459 TVVDLEVCICNDG-STDNTLEVIN-----KLYGNPNRVRIMS-----KENGGIASASN 505
Db 397 AGFYTFRRSGW-----DKRATVMVLKASPPGEHAQPDNGTFLFKGRNFTPDA 445
Db 506 AA-VSFAKGYIGQLDSDDDYLEPDADVELCLK-----EF--LKDKTLACVYTTNRNVPDG 557

Qy 446 GVFV--YSGDE-----AIMKLRWYRQTRIHSIL-TLDNQNMVITKARQN 487
Db 558 SLIANGYNWPEFSREKLTATTAMIAHFRMFTIRAWHLTDGFENENIENAVDYMFLKJSEVG 617
Qy 488 KWETGNL---DVLTYNPSYNLDHORSVLFINKKYFLVIDRAIGEATGNLGVHW--- 540
Db 618 KPHLKNKICYNRVLHGDNSTIKKLGIQK-----KNHFVVVNGSLNRQ---GLNYNYND 667
Qy 541 ---QLKEDSNPVDKTKNRVYTTYRDGNNLMIOSLNADRTSLNEBEGKVSYYVYNKELKRP 597
Db 668 KFDLDDESRYIENKT---AEYQEBMDML-----KDLK-- 697
Qy 598 AFVFEKPKKNAQTQNFYSIYVP 619
Db 698 -LIQNKDAKIA-----VSIFYP 713

RESULT 14
US-10-184-485-3
; Sequence 3, Application US/10184485
; Publication No. US2003007763A1
; GENERAL INFORMATION:
; APPLICANT: THE BOARD OF REGENTS OF THE UNIVERSITY OF OKLAHOMA
; TITLE OF INVENTION: POLYMER GRAFTING BY POLYSACCHARIDE SYNTHASES
; FILE REFERENCE: 5820.551
; CURRENT APPLICATION NUMBER: US/10/184.485
; CURRENT FILING DATE: 2002-06-26
; PRIOR APPLICATION NUMBER: US/09/437,277
; PRIOR FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: WordPerfect 8.0 (saved in ASCII format)
; SEQ ID NO 3
; LENGTH: 965
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-10-184-485-3

Query Match 3.0%; Score 105; DB 15; Length 965;
Best Local Similarity 17.7%; Pred. No. 1.6;
Matches 131; Conservative 115; Mismatches 238; Indels 258; Gaps 38;

Qy 13 AVIALSSGNILAQSSITRKDF-----DHIN-----LEYSGLEKVNKAVAAG 54
Db 95 SITGKKSNAEIRKVELVPKDFPKDLVLAFLPDHVNDFTWYKNRKSIGIKVPVKNIGLS 154
Qy 55 -----NY-----DAAKALLAYYREKSKAREPDFSNAEK 83
Db 155 IIIPTFNRSRLDITLACLNVQKNTYFVVVAVDGGSKENLLTIVQK-----YEQ 204
Qy 84 PADIRQPIDKVTREMAKALVHQPQPHKGYG-----FDYKGDINQWMP- 128
Db 205 KLDIK-----YVRQKDYGYQLCAVRNLGLRTAKYDFVSIILCDMAPO 246
Qy 129 -----VKDNEV-----RWQLHRVKKWQAMALVYHATGDEKYAREWYQYSDW 170
Db 247 QLWVHSYLTLELNDIVLIGPRKYVDTHNTAQFL-----NDPYLIESLPETA-- 296
Qy 171 ARKNPLGLSQDNKDFVWR--PLEVSDRVQSLPPTFSLFVNSPAPTPAFLEFNLNS--YH 225
Db 297 TNNPSTISKGNISLDWRLEHFKTDNLRLCDSPPRYFV---AGNVAFSKEWLNKVGWFD 353
Qy 226 QADYLSTHYAEOGNHRLFEAQ---RNLFAGVSP---PEFKDSPRWQTGISVLNTEIKK 279
Db 354 EEFNHWGGEDEVFG-YRLFAGKCFRVIDGGMALHQEPGKENETEBEAGKSITLKIYKE 412
Qy 280 QV---YADGMQFELSPIYHVAADIFLKAYGSAKRVNLEKFFPOSYYQTVENMMALISI 336
Db 413 KVPYIYRKLPIEDSHIRIPLVSIYPAYNCA-----NYIQ---RCVDSALNQ 458
Qy 337 SLDPYNTPMFGDSWITDKNFRMAQFASWARVFPANQAIKYFATDGKQKGAENFLSKALSN 396
Db 459 TVVDLEVCICNDG-STDNTLEVIN-----KLYGNPNRVRIMS-----KENGGIASASN 505

This Page Blank (uspto)